Journal of Business Law

Issue 5 2017

Table of Contents

This journal should be cited as [2017] J.B.L.

The *Journal of Business Law* is published by Thomson Reuters (Professional) UK Limited, trading as Sweet & Maxwell. Registered in England & Wales, Company No.1679046. Registered Office and address for service: 5 Canada Square, Canary Wharf, London, E14 5AQ.

For further information on our products and services, visit *http://www.sweetandmaxwell.co.uk*

Computerset by Sweet & Maxwell. Printed and bound in Great Britain by Hobbs the Printers Ltd, Totton, Hampshire.
No natural forests were destroyed to make this product: only farmed timber was used and replanted.

ISSN 0021–9460

LEGAL TAXONOMY
FROM SWEET & MAXWELL

Each article and case commentary in this issue has been allocated keywords from the Legal Taxonomy utilised by Sweet & Maxwell to provide a standardised way of describing legal concepts. These keywords are identical to those used in Westlaw UK and have been used for many years in other publications such as Legal Journals Index. The keywords provide a means of identifying similar concepts in other Sweet & Maxwell publications and online services to which keywords from the Legal Taxonomy have been applied. Keywords follow the taxonomy logo at the beginning of each item. The index has also been prepared using Sweet & Maxwell's Legal Taxonomy. Main index entries conform to keywords provided by the Legal Taxonomy except where references to specific documents or non-standard terms (denoted by quotation marks) have been included. Readers may find some minor differences between terms used in the text and those which appear in the index. Please send any suggestions to *sweetandmaxwell.taxonomy@tr.com*.

Orders by email to: *TRLUKI.orders@thomsonreuters.com*.

Copies of articles from the *Journal of Business Law*, and other articles, cases and related materials, can be obtained from DocDel at Sweet & Maxwell's Yorkshire office. Current rates are: £7.50 + copyright charge + VAT per item for orders by post, DX and email. Fax delivery is guaranteed within 15 minutes of request and is charged at an additional £1.25 per page (£2.35 per page outside the UK). For full details, and how to order, please contact DocDel on
Tel: 01422 888 019. Fax: 01422 888 001.
Email: *trluki.admincentral@thomsonreuters.com*.
Go to: *http://www.sweetandmaxwell.co.uk/our-businesses/docdel.aspx*.
Please note that all other enquiries should be directed to Customer Support (Email: TRLUKI.cs@thomsonreuters.com; Tel: 0345 600 9355).

Brexit and Financial Services: A Tentative Analysis of Possible Exit Scenarios

Maziar Peihani*
Centre for International Governance Innovation

keywords to be inserted by the indexer

Introduction

On 23 June 2016, the British public voted to leave the EU by a margin of 52 to 48 per cent.[1] Although the outcome of the referendum is not binding, the UK Government has stressed that "Brexit means Brexit" and it will therefore respect the will of the majority.[2] In fact, the UK Government has introduced the European Union (Notification of Withdrawal) Bill and intended to trigger the withdrawal process by the end of March 2017.[3] The withdrawal from the EU is governed by art.50 of the Lisbon Treaty, whereby any state that seeks to leave the EU should serve notice to the European Council.[4] The EU and the state in question will then negotiate and conclude an agreement which sets out the terms of divorce. The EU treaties will stop applying to the exiting Member State from the time the withdrawal agreement comes into force or, failing that, two years after the notice is filed, unless the European Council extends this period.[5]

Brexit is certainly one of the most remarkable events in recent European history, marking the first time that an EU Member State has decided to leave the bloc. In addition to important implications for European integration, Brexit has created significant uncertainty about a vast range of issues, including the rules that govern the UK's trade with the EU and the rest of the world, the rights of EU workers in the UK and vice versa, and the resilience of the UK economy in coming years.[6]

* Maziar Peihani is a post-doctoral fellow with the International Law Research Program.

1 A. Hunt and B. Wheeler, "Brexit: All You Need to Know about the UK Leaving the EU-What Was the Breakdown Across the UK?" (last updated on 6 March 2017), *BBC News*, *http://www.bbc.com/news/uk-politics-32810887* [Accessed 16 May 2017].

2 "Theresa May's Brexit Speech in Full" (17 January 2017), *The Telegraph*, *http://www.telegraph.co.uk/news/2017/01/17/theresa-mays-brexit-speech-full/* [Accessed 16 May 2017].

3 UK Parliament, European Union (Notification of Withdrawal) Bill 2016–17 (accessed on 6 March 2017) *http://services.parliament.uk/bills/2016-17/europeanunionnotificationofwithdrawal.html* [Accessed 16 May 2017].

4 Treaty of Lisbon amending the Treaty on European Union and the Treaty establishing the European Community, signed at Lisbon [2007] OJ C306 art.50 (Treaty of Lisbon).

5 Treaty of Lisbon [2007] OJ C306 art.50.

6 See, e.g., B. S. Bernanke, "Economic Implications of Brexit" (28 June 2016), *Brookings Institution*, *https://www.brookings.edu/blog/ben-bernanke/2016/06/28/economic-implications-of-brexit/*; Markus Gehring, "What Are the Legal and Constitutional Impacts of Brexit?" (30 June 2016), Centre for International Governance Innovation (CIGI), *https://www.cigionline.org/articles/what-are-legal-and-constitutional-impacts-brexit*; A. Beattie, "Vision of Post-Brexit Trade Puzzles Experts" (23 January 2017), *Financial Times*, *https://www.ft.com/content/0d48300a-de3b-11e6-86ac-f253db7791c6*; S. Donnan, "Trade Truths for Theresa May's Vision of a 'Global Britain'" (18 January 2017), *Financial Times*, *https://www.ft.com/content/65322fbc-dd34-11e6-9d7c-be108f1c1dce*; Bank of England, "Monetary Policy Summary Publication" (3 November 2016), p.3, *http://www.bankofengland.co.uk/publications/minutes*

Such uncertainty is perhaps most profound with respect to financial services, a globally oriented industry greatly reliant on unfettered access to the European markets and infrastructure. EU financial exports account for 39 per cent of the total EU financial services gross value added, with 22 per cent of such trade occurring within the EU. The UK is greatly involved in the intra-EU trade flows.[7] For example, it accounts for 78 per cent of foreign exchange trading, 74 per cent of interest rate derivatives, and 50 per cent of fund management in Europe.[8] Once the UK has left the EU, UK financial intermediaries and international firms that have chosen London as their European base will lose their passports to conduct cross-border trade and serve clients in the 27 EU Member States.

The key question that therefore arises is how to govern future relations between the UK and the EU in the realm of financial services. This question is the key focus of this article, which discusses three primary scenarios that can govern the parties' future relationship: European Economic Area (EEA membership); third-country equivalence; and a bespoke agreement. The article carefully assesses each option, the opportunities and challenges they present, and the key legal and regulatory issues they give rise to. It is argued that EEA membership offers the greatest access to the single market, posing the least disruption to the smooth functioning of financial intermediation across Europe. At the same time, however, it poses the greatest political challenge for the UK Government, which has chosen a Brexit strategy that focuses on gaining back full control of immigration and staving off the jurisdiction of the Court of Justice of the EU (CJEU).

The discussion on equivalence suggests that it is a relatively new regime only available under certain EU legislations. Not only does equivalence exclude important areas of financial activity, but its assessment and determination is a process administered by the EU Commission which can be influenced by politics. The Commission can grant access on a partial or provisional basis and withdraw it altogether. Moreover, maintaining equivalence can be challenging as the two regimes will inevitably grow more divergent over time. Finally, the article discusses the concept of a bespoke arrangement and the agreements that the EU has concluded with third countries such as Switzerland and Canada. It suggests that the existing arrangements do not offer good models for post-Brexit negotiations as they do not match the UK's trade profile, take significant time to conclude, and offer relatively narrow access to the single market. The article also questions whether the UK can obtain preferential regulatory equivalence and access in a bespoke arrangement when it is adamant to leave the single market and the CJEU's jurisdiction.

The article starts with an overview of the UK's financial sector, and the freedoms and passport rights that UK financial institutions currently enjoy. It maps the passporting into primary types of financial intermediation such as banking, insurance and asset management, and discusses how significant it is for each sector. The article then delves into possible options that can govern future relations between

/Documents/mpc/mps/2016/mpsnov.pdf; Bank of England , "Bank of England Inflation Report Q&A 2nd February 2017" (2 February 2017), pp.13–15, *http://www.bankofengland.co.uk/publications/Documents/inflationreport/2017/conf020217.pdf* [All accessed 16 May 2017].

[7] PWC, "Planning for Brexit: Operational Impacts on Wholesale Banking and Capital Markets in Europe" (January 2017), p.2, *Association for Financial Markets in Europe*, *https://www.afme.eu/globalassets/downloads/publications/afme-pwc-planning-for-brexit.pdf* [Accessed 16 May 2017].

[8] PWC, "Planning for Brexit" (January 2017), p.2, *Association for Financial Markets in Europe*, *https://www.afme.eu/globalassets/downloads/publications/afme-pwc-planning-for-brexit.pdf* [Accessed 16 May 2017].

the UK and the EU. It concludes the discussion by reflecting upon the path forward, as well as outstanding transitional issues. It stresses that it is in both parties' interest to find a workable solution to safeguard the valuable elements of the single market, such as passporting. However, it also warns that the UK's Brexit strategy as it currently stands is founded on unrealistic and irreconcilable objectives, which risks driving the UK's economy off the cliff.

A primer on the UK's financial sector, single market and passporting

The UK, and the City of London in particular, has been a leading financial centre for centuries. Finance constitutes one of the most important areas of economic activity in the UK. The Office for National Statistics estimates the financial sector output to be 8 per cent of the UK's national output.[9] Some have argued that if relevant business services are included, this number can be as high as 1 per cent.[10] The financial sector is also an important taxpayer. In the year up to 31 March 2016, the sector's tax contributions reached a record high of £66 billion.[11] The country's trade surplus in financial services was £63 billion in 2015, which is larger than the combined surpluses of the next three leading competitors, namely the US, Switzerland, and Luxembourg.[12]

The UK has one of the largest financial systems in the world. Standing at about £20 trillion, the sum of financial assets owned by UK financial institutions, excluding the Bank of England, is over 10 times the UK's annual GDP.[13] Nearly a fifth of global banking activity is booked in the UK and around half of the world's largest financial institutions, including banks, insurers, and asset managers, have their European headquarters in the UK.[14] Four UK banks—HSBC, Barclays, RBS and Standard Chartered—have been designated as Global Systemically Important Banks.[15] The UK's insurance sector is the largest in Europe and third largest in the world.[16] In addition, the UK hosts the largest wealth management industry as well as many of the important equity trading platforms in Europe.[17] Over half of the euro zone firms raise capital in London in the form of equity or debt. Finally, the UK is also a hub for securities and derivatives trading, hosting two of the largest Central Clearing Counterparties (CCP) in the world.[18]

[9] G. Tyler, "Financial Services: Contribution to the UK Economy" (House of Commons Library, SN/EP/06193, 26 February 2015), p.1, *http://researchbriefings.parliament.uk/ResearchBriefing/Summary/SN06193* [Accessed 16 May 2017].

[10] A. Armstrong, "EU Membership, Financial Services and Stability" (2016) 236 *National Institute Economic Review* 31, 31.

[11] City of London, "Total Tax Contribution of UK Financial Services" (December 2016), p.5, *https://www.cityoflondon.gov.uk/business/economic-research-and-information/research-publications/Documents/Research%202016/total-tax-report-2016.pdf* [Accessed 16 May 2017].

[12] TheCityUK, "Key Facts about UK Financial and Related Professional Services" (March 2016), p.9, *https://www.thecityuk.com/assets/2016/Reports-PDF/Key-facts-about-UK-financial-and-related-professional-services-2016.pdf* [Accessed 16 May 2017].

[13] IMF, "United Kingdom—Financial Sector Assessment Program: Financial System Stability Assessment" (IMF Country Report No.16/167, June 2016), p.9.

[14] IMF, "United Kingdom—Financial Sector Assessment Program" (June 2016), p.9.

[15] Financial Stability Board, "2016 list of global systemically important banks (G-SIBs)" (November 2016), p.3, *http://www.fsb.org/wp-content/uploads/2016-list-of-global-systemically-important-banks-G-SIBs.pdf* [Accessed 16 May 2017].

[16] IMF, "United Kingdom—Financial Sector Assessment Program" (June 2016), p.9.

[17] IMF, "United Kingdom—Financial Sector Assessment Program" (June 2016), p.9.

[18] IMF, "United Kingdom—Financial Sector Assessment Program" (June 2016), p.9.

As a member of the EU the UK enjoys freedom of movement of capital as well as freedom of financial services.[19] Free movement of capital allows UK households and firms to borrow and invest abroad and make cross-border payments.[20] The free movement of financial services has two dimensions: the freedom to provide financial services and the freedom of establishment. While both freedoms date back to the Treaty of Rome (1957),[21] a key development in the free movement of financial services came in the 1990s when the EU introduced the passporting regime. The concept of passporting relies on mutual recognition of prudential measures by member counties, coupled with minimum EU standards. It seeks to minimise legal, regulatory and operational barriers to cross-border provision of financial services in the European Economic Area (EEA). A firm that is authorised by a regulator in one Member State is therefore allowed to carry out the same permitted activities in another Member State. It can do so by either directly providing cross-border financial services, or by setting up a branch in the other Member State.[22]

Passporting has been introduced through several pieces of EU legislation for various financial activities and services.[23] Table 1 shows the EU legislations that provide for passporting rights in financial services. It also shows the number of inbound and outbound passports that have been issued under each piece of legislation. An outbound passport refers to a passport that has been issued by the UK's competent authorities, namely the Financial Conduct Authority (FCA) and the Prudential Regulation Authority (PRA). An inbound passport means a passport issued by an EU or EEA competent authority firm in their states, enabling them to do business in the UK.

Table 1: Number of firms with at least one passport under each directive
Source: Financial Conduct Authority, August 2016

Directive	Outbound	Inbound
Alternative Investment Fund Managers Directive (AIFMD)	212	45
Insurance Mediation Directive (IMD)	2,758	5,727
Markets in Financial Instruments Directive (MiFID)	2,250	988
Mortgage Credit Directive (MCD)	12	0
Payment Services Directive (PSD)	284	115

[19] Treaty on the Functioning of the European Union, 25 March 1957, Title IV, Free Movement of Persons, Services and Capital (entered into force 1 January 1958) (TFEU).

[20] Bank of England, "EU Membership and the Bank of England" (October 2015), p.20, *http://www.bankofengland.co.uk/publications/Documents/speeches/2015/euboe211015.pdf* [Accessed 16 May 2017].

[21] Treaty of Rome, 25 March 1957, Title I, Free Movement of Goods and Title III, Free Movement of Persons, Services and Capital.

[22] Bank of England, "EU Membership and the Bank of England" (October 2015), p.24, *http://www.bankofengland.co.uk/publications/Documents/speeches/2015/euboe211015.pdf* [Accessed 16 May 2017].

[23] "EU Membership and the Bank of England" (October 2015), p.24, *http://www.bankofengland.co.uk/publications/Documents/speeches/2015/euboe211015.pdf* [Accessed 16 May 2017].

Directive	Outbound	Inbound
Undertakings for the Collective Investment in Transferable Securities Directive (UCITS)	32	94
Electronic Money Directive	66	27
Capital Requirements Directive IV (CRD IV)	102	552
Solvency II Directive	220	726

[24]From the banking perspective, the CRD IV and MiFID are particularly important. The CRD allows the banks based in the UK to lend directly to corporations based anywhere in the EU or conduct business through the establishment of branches.[25] These branches remain under the supervisory authority of the home country, namely the UK. Important business activities covered by the CRD IV's passport are: deposit taking, lending brokering, payment services, securities issuance and portfolio management.[26] The passport under the MiFID operates similar to the CRD IV and covers the following business activities: executing orders for clients as well as trading on one's own account, investment advice, underwriting, foreign exchange services and portfolio management.[27] Industry reports indicate that UK banks can lose up to 20 per cent of their revenue if they lose access to the EU's passport.[28]

For the investment fund industry, including asset managers, money market funds, hedge funds, private equity and venture capital, the passport is largely introduced through the AIFMD and the UCITS.[29] Together, these directives allow investment funds to market their products in any EU Member state. More specifically, these legislations provide for two types of passporting: (1) management passport, allowing asset managers to provide services to funds domiciled in EU countries; and (2) marketing passport, allowing EU domiciled funds to be managed in the UK but marketed and sold across the EU.[30]

The passport regime for investment funds is less complete than for banking and therefore of less value. While services can be offered in any EU Member State,

[24] Andrew Bailey, Letter from the Financial Conduct Authority to Andrew Tyrie, Chairman of the Treasury Committee House of Commons (17 August 2016), *https://www.parliament.uk/documents/commons-committees/treasury/Correspondence/AJB-to-Andrew-Tyrie-Passporting.PDF* [Accessed 16 May 2017].

[25] Directive 2013/36 on access to the activity of credit institutions and the prudential supervision of credit institutions and investment firms, amending Directive 2002/87 and repealing Directives 2006/48 and 2006/49 [2013] OJ L176/338, arts 33–34 (CRD IV); Directive 2004/39 on markets in financial instruments amending Directives 85/611 and 93/6 and Directive 2000/12 and repealing Directive 93/22 [2004] OJ L145 arts 32–34 (MifID).

[26] CRD IV, Annex I, List of Activities Subject to Mutual Recognition.

[27] MiFID, Annex I, List of Services and Activities and Financial Instruments, ss.A and B.

[28] For example, a recent Oliver Wyman study suggests that UK banking sector's total revenue in 2015 was between £108 billion and £117. Of this, between £23 billion and £27 billion were international and wholesale business related to the EU. It therefore estimates that 21% to 23% of UK banks' revenue can be affected by loss of passporting. See O. Wyman, "The Impact of the UK's Exit from the EU on the UK-Based Financial Services Sector" (2016), p.6, *http://www.oliverwyman.com/content/dam/oliver-wyman/global/en/2016/oct/Brexit_POV.PDF* [Accessed 16 May 2017].

[29] Directive 2011/61 on Alternative Investment Fund Managers and amending Directives 2003/41 and 2009/65 and Regulations 1060/2009 and 1095/2010 [2011] OJ L174 arts 31 and 32 (AIFM); Directive 2009/65 on the coordination of laws, regulations and administrative provisions relating to undertakings for collective investment in transferable securities [2009] OJ L302 art.16 (UCITS).

[30] Open Europe, "How the UK's Financial Services Sector Can Continue Thriving after Brexit" (October 2016), p.23, *http://openeurope.org.uk/intelligence/britain-and-the-eu/how-the-uks-financial-services-sector-can-continue-thriving-after-brexit/* [Accessed 16 May 2017].

the actual marketing of funds is still subject to national regulation.[31] For example, some national regulators charge fees on asset managers that are based in other countries. In addition, there are important divergences between national regimes on what constitutes a "professional investor" or "marketing".[32] Nonetheless, European business is still of significance to the UK's investment fund industry. The latest study published by the Investment Association suggests that UK firms managed £1.2 trillion in assets for European clients in 2015.[33] The loss of access to the EU single market can therefore have significant implications for the UK investment funds industry.

For the insurance industry, the important legislations are Solvency II and the MiFID. The Insurance Mediation Directive II, which will be replaced by the Insurance Distribution Directive as of February 2018, also has some significance as it governs the sale and disclosure of insurance products.[34] It has been argued that the UK insurers do not use passporting as widely as the banks.[35] Open Europe, for example, estimates that 87 per cent of the cross-border insurance business is conducted through subsidiaries and only 13 per cent is done through branches.[36] One key reason behind this trend is that insurance firms prefer to keep risks isolated in separately capitalised subsidiaries. This business strategy is particularly common for retail insurance where a presence on the ground and local knowledge are particularly important. Caution must be, however, taken in interpreting the numbers cited above. Lloyd's London, which retains a significant share of the insurance market, relies significantly on passporting to provide underwriting services either directly or through branches in other Member States. In 2015, the EU/EEA accounted for £2.9 billion or 11 per cent of Lloyd's gross written premium.[37] As Huw Evans, the director-general of the Association of British Insurers, points out, the significance of this number is understood far better when it is noted that Lloyd's annual revenue in 2015 was £27 billion,[38] amounting to 64 to 69 per cent of total UK's insurance revenue.[39] Passporting is, therefore, far more important to the UK's insurance industry than is often noted in the quantitative studies on Brexit.

[31] Open Europe, "How the UK's Financial Services Sector Can Continue Thriving after Brexit" (October 2016), p.24, *http://openeurope.org.uk/intelligence/britain-and-the-eu/how-the-uks-financial-services-sector-can-continue-thriving-after-brexit/* [Accessed 16 May 2017].

[32] European Markets and Securities Authority (ESMA), "Opinion to the European Parliament, Council and Commission, and Responses to the Call for evidence on the Functioning of the AIFMD EU Passport and of the National Private Placement Regimes" (30 July 2015), p.16, *http://www.esma.europa.eu/sites/default/files/library/2015/11/2015-1235_opinion_to_ep-council-com_on_aifmd_passport_for_publication.pdf* [Accessed 16 May 2017].

[33] The Investment Association, "Asset Management in the UK 2015–2016" (September 2016), p.16, *http://www.theinvestmentassociation.org/assets/files/research/2016/20160929-amsfullreport.pdf* [Accessed 16 May 2017].

[34] Directive 2002/92 on insurance mediation [2002] OJ L009; Directive 2016/97 on insurance distribution (recast) Text with EEA relevance [2016] OJ L26.

[35] D. Schoenmaker, "The UK Financial Sector and EU Integration after Brexit: The Issue of Passporting" (October 2016), p.5, *SSRN*, *https://ssrn.com/abstract=2844253* [Accessed 16 May 2017].

[36] Open Europe, "How the UK's Financial Services Sector Can Continue Thriving after Brexit" (October 2016), p.27, *http://openeurope.org.uk/intelligence/britain-and-the-eu/how-the-uks-financial-services-sector-can-continue-thriving-after-brexit/* [Accessed 16 May 2017].

[37] M. Faulkner, "Lloyd's In Talks to Maintain EU Licenses on Brexit" (1 July 2016), Lloyd's List, *https://www.lloydslist.com/ll/sector/insurance/article529493.ece* [Accessed 16 May 2017].

[38] Select Committee on the European Union, Financial Affairs Sub-Committee, "Corrected Oral Evidence: Brexit and Financial Services in the UK" (12 October 2016), Q51, *http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/eu-financial-affairs-subcommittee/brexit-financial-services/oral/41228.html* [Accessed 16 May 2017].

[39] These figures are author's calculations based on the UK's total insurance revenue estimated by the European Parliament in December 2016. See European Parliament, "Brexit: the United-Kingdom and EU Financial Services" (Briefing, PE 587.3849, December 2016), p.1, *http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/587384/IPOL_BRI(2016)587384_EN.pdf* [Accessed 16 May 2017].

While there is a clear understanding that maintaining passporting rights is crucial to the UK's financial services, assessing the economic costs of losing such rights has proved difficult. A quantitative study by Oliver Wyman suggests that nearly a quarter of revenues in banking and asset management, and nearly half of the revenues in market infrastructure and other areas, are related to the EU.[40] Wyman estimates that a low access scenario can result in total revenue loss of about £18–20 billion.[41] Recent House of Lords hearings on Brexit, however, suggest that it is difficult to read much into these numbers.[42] Neither the public authorities nor the financial industry seem to understand yet how passporting maps into the business structure and the operation of financial institutions.[43] Indeed, it is difficult if not impossible to isolate various products and services and then quantify the impact of Brexit under different scenarios. As Katherine Braddic, director of financial services, HM Treasury, notes, the UK's financial sector is a complex and interconnected eco-system which has long defied quantitative analysis.[44] While explanations can be offered on how financial institutions and markets operate, translating those explanations into impact scenarios remains difficult.[45] Consequently, the impact of Brexit on the UK's financial ecosystem remains quite unclear, especially when no coherent information is yet available on the UK Government's strategy or the direction of future negotiations.

Governing the UK-EU future relations—a survey of primary options

The key question that has arisen since the Leave vote in June's referendum is how future relations between the UK and the EU can be governed. This question overlaps with many key areas of social and economic activities, including investment, trade and environment, though the focus of the current article is primarily on financial services. While there are infinite Brexit scenarios, three options seem to be most relevant when it comes to governing the future relations of both parties: a European Economic Area (EEA) membership, equivalence or a bespoke arrangement. The first option offers the greatest access to the single market and the least disruption to the smooth functioning of financial institutions. The second option relies on specific EU legislations that allow third countries to gain access to the single market on a case-by-case basis. Finally, the third option is a bespoke free trade agreement which seeks to ensure that the UK maintains access to the single market.

[40] Wyman, "The Impact of the UK's Exit from the EU on the UK-Based Financial Services Sector" (2016), p.6, *http://www.oliverwyman.com/content/dam/oliver-wyman/global/en/2016/oct/Brexit_POV.PDF* [Accessed 16 May 2017].

[41] The low access scenario means that the UK becomes a third country without receiving equivalence or preferential access on a bilateral basis. Wyman, "The Impact of the UK's Exit from the EU on the UK-Based Financial Services Sector" (2016), p.14, *http://www.oliverwyman.com/content/dam/oliver-wyman/global/en/2016/oct/Brexit_POV.PDF* [Accessed 16 May 2017].

[42] House of Lords European Union Committee, "Brexit: Financial services" (9th Report of Session 2016–17, December 2016), p.13.

[43] House of Lords European Union Committee, "Brexit: Financial services" (December 2016), p.15.

[44] Select Committee on the European Union, Financial Affairs Sub-Committee, "Corrected Oral Evidence: Brexit and Financial Services in the UK" (19 October 2016), Q58, *http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/eu-financial-affairs-subcommittee/brexit-financial-services/oral/41565.html* [Accessed 16 May 2017].

[45] Select Committee on the European Union, "Corrected Oral Evidence: Brexit and Financial Services in the UK" (19 October 2016), Q58, *http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/eu-financial-affairs-subcommittee/brexit-financial-services/oral/41565.html* [Accessed 16 May 2017].

At the time of writing this article, the third option seems the most favourable to the British Government. The Prime Minister's speeches and a recent White Paper on Brexit have repeatedly called for a bold and ambitious free trade agreement.[46] In spite of this preference, the other two options still remain relevant and worthy of analysis. This is particularly the case as the Government's proposed bespoke arrangement has elements of both the EEA membership and equivalence. The Government not only seeks to conclude a free trade agreement but also to maintain an EEA-type access to the single market and ensure the continued equivalence of legal and regulatory regimes with the EU.[47] It is also important to be mindful of the UK's internal politics, and that Brexit preferences can change over time. The UK Government is just one important actor in the Brexit process. The business community, those who voted to remain in the EU and the UK Parliament can also influence the path to Brexit to different degrees. The recent UK Supreme Court ruling on art.50 is particularly noteworthy in this respect as it confirms the principle of parliamentary sovereignty and the Government's duty to consult Parliament over Brexit.[48] Parliamentary debate and scrutiny can therefore pressure the Government to re-evaluate and change its Brexit strategies and objectives. The following sections will therefore discuss the legal and regulatory issues that arise under all three Brexit options.

European Economic Area (EEA) membership

The EEA refers to a single market established between EU Member States on one hand and the three members of the European Free Trade Association (EFTA), namely, Iceland, Lichtenstein and Norway.[49] These countries are commonly referred to as EEAEFTA members, and their relationship with the EU is underpinned by the European Economic Area Agreement.[50] Switzerland, another EFTA member, has its own agreement with the EU, which will be discussed at a later stage in this article. The EEA Agreement grants virtually full access to the single market and provides for the four freedoms, namely free movement of goods, services, capital and labour, in the same way as they are applicable within the EU.[51] The EEA, however, is not a customs union and excludes agriculture, fisheries, common trade policy and foreign policy.

The EEA Agreement provides for the simultaneous application of EU rules on the internal market, state aid and competition within the EAA.[52] This is meant to ensure that the rules of the single market remain homogenous. The EU legislations are imported as an annex to the EEA Agreement, and their implementation is

[46] "Theresa May's Brexit Speech in Full" (17 January 2017), *The Telegraph*, *http://www.telegraph.co.uk/news/2017/01/17/theresa-mays-brexit-speech-full/* [Accessed 16 May 2017]; HM Government, "The United Kingdom's EXIT from and New Partnership with the European Union" (February 2017), p.35, *https://www.gov.uk/government/publications/the-united-kingdoms-exit-from-and-new-partnership-with-the-european-union-white-paper* [Accessed 16 May 2017] (Brexit White Paper).

[47] Brexit White Paper (February 2017), pp.42–43, *https://www.gov.uk/government/publications/the-united-kingdoms-exit-from-and-new-partnership-with-the-european-union-white-paper* [Accessed 16 May 2017].

[48] *R. (on the application of Miller) v Secretary of State for Exiting the European Union* [2017] UKSC 5; [2017] 2 W.L.R. 583 at [110], [124].

[49] "European Economic Area", *http://www.efta.int/eea* [Accessed 16 May 2017].

[50] Agreement on the European Economic Area [1994] OJ L1 (EEA Agreement).

[51] EEA Agreement art.1.

[52] EEA Agreement art.102.

monitored by the EFTA Surveillance Authority.[53] The decisions of the EFTA Surveillance Authority can be appealed to the EFTA Court. The EFTA Court can also hear actions against the Surveillance Authority, as well as disputes between the EFTA Member States.[54] The EFTA Court consists of three judges and its decisions are binding upon the EFTA Member States. The EFTA Court takes into account the case law of the Court of Justice of the EU (CJEU) when interpreting the EEA agreement, as well as EU treaties and legislations.[55]

If the UK decides to join the EEA, disruptions to the financial sector will be minimised. UK financial institutions will continue to benefit from the passport regime by operating through branches or providing direct cross-border services. Non-EU financial institutions can also maintain their commercial presence in the UK and do not need to move their headquarters to other EU/EEA countries as the UK would remain the gateway to the EU markets. How can the UK join the EEA? Some have suggested that the UK will be an EAA member by default once it has left the EU. This proposition, however, seems implausible, as the EEA Agreement applies only between the EU members and EFTA states. Once the withdrawal procedure under art.50 is complete,[56] the EU treaties cease to apply to the UK.[57] This means that the UK will lose its EU member status and will also be automatically excluded from the EEA Agreement. In order to join the EEA, the UK needs to follow the procedures under art.108 of the EEA Agreement. Under this provision, any European state that seeks to become an EEA party should submit its application to the EEA Council, the highest decision-making authority in the EEA.[58] The terms and conditions of the accession need to be agreed upon by all "Contracting Parties" to the EEA Agreement, namely the EU and EFTA Member States. The agreement should then be submitted for ratification to the contracting parties under their domestic laws.[59]

However, while offering significant benefits, EEA membership comes with important costs and challenges. First, the UK is expected to make a contribution to the EU's budget in return for access to the single market and other EU programmes. Based on current estimates, this contribution will not significantly differ from what the UK is currently paying. For example, Open Europe estimates that Norway's per capita net contribution to the EU budget in 2015 was £100.03. This number is as high as 88 per cent of the UK's contribution, £113 per capita, in the same period.[60] Secondly, in return for access to the single market, the UK needs to respect the four freedoms and submit to the jurisdiction of the EFTA Court, which closely follows the CJEU's precedent. The EEA Agreement allows the Member States to unilaterally adopt safeguard measures in case of "serious

[53] EEA Agreement art 109.
[54] EEA Agreement art 108.
[55] Agreement between the EFTA States on the establishment of a surveillance authority and a Court of Justice [1994] OJ L344 art.3.
[56] Treaty of Lisbon art.50.
[57] Treaty of Lisbon art.2(c).
[58] EEA Agreement art.90. The EEA Council consists of the EFTA governments, members of the EU Council and representatives from the European Commission and European External Action Service.
[59] EEA Agreement art.126.
[60] S. Booth, "As the UK Searches for a Post-Brexit Plan, is the EEA a Viable Option?" (4 August 2016), *Open Europe*, *http://openeurope.org.uk/intelligence/britain-and-the-eu/as-the-uk-searches-for-a-post-brexit-plan-is-the-eea-a-viable-option/*; S. Ashworth-Hayes, "Norwegians Pay About As Much As Brits to Access EU" (2 November 2016), *https://infacts.org/norwegians-pay-same-brits-eu-access/* [Both accessed 16 May 2017].

economic, societal or environmental difficulties”.[61] As the language of art.112 indicates, however, these measures can only be adopted in exceptional circumstances and must be temporary and limited in scope. Moreover, the adoption of such measures entitles other contracting parties to take “proportionate rebalancing measures,” which can mean restricting access to the single market.[62] Owing to such restrictions, safeguard measures have been used in a limited way. Norway, for example, has never used safeguard measures. While Lichtenstein has adopted safeguards to restrict the free movement of people, such measures have been reached by way of an agreement with the EU which Lichtenstein cannot amend unilaterally. More importantly, Lichtenstein’s safeguards are very limited in scope and can be explained by its unique circumstances: a population of around 37,000 and a territory of around 160 square kilometers.[63] The final challenge has to do with the UK’s influence over the EU rule-making process. As mentioned earlier, EU legislations are transposed into the EEA legal order. Member States are expected to follow EU rules on core freedoms, state aid, competition and so on.[64] While the EFTA states can express their views on the proposed EU rules, they can have no vote on what is decided and can have only limited influence over the rule-making process.[65]

These challenges are of varying magnitude. For example, the issue of financial commitment seems to be of less concern given that the UK Government has recently expressed readiness to make some form of contribution to the EU budget in return for access to the internal market. Similarly, the UK’s influence over the rule-making process can also be reinforced by greater emphasis on the EEA Agreement provisions which seek to facilitate timely input from Member States in the rule-making process.[66] Given the UK’s significance as a major financial jurisdiction, it is likely that its views will have more influence over the EU’s rule-making process on financial services than other EFTA states. The second problem, however, seems to pose the greatest challenge to the viability of the EEA option. The UK Prime Minister has expressly indicated that her country is not leaving the EU to give up control of immigration or return to the jurisdiction of the CJEU.[67] If the UK seeks to resume full sovereignty over immigration and rule of law, EEA membership cannot be pursued as a realistic option.

Third-country equivalence

Certain EU financial legislations allow a third country to gain access to the single market provided that its legal and regulatory frameworks are recognised as equivalent to the EU’s. This recognition enables the EU to rely on the foreign firms’ compliance with an equivalent regulatory framework, allowing them access

[61] EEA Agreement art.112.

[62] EEA Agreement art.114.

[63] Booth, “As the UK Searches for a Post-Brexit Plan, is the EEA a Viable Option?” (4 August 2016), *Open Europe*, *http://openeurope.org.uk/intelligence/britain-and-the-eu/as-the-uk-searches-for-a-post-brexit-plan-is-the-eea-a-viable-option/* [Accessed 16 May 2017].

[64] EEA Agreement Pt IV, Chs I and II.

[65] J.-C. Piris, “If the UK Votes to Leave: The Seven Alternatives to EU Membership” (January 2016), p.6, *Centre for European Reform*, *https://www.cer.org.uk/sites/default/files/pb_piris_brexit_12jan16.pdf*.

[66] EEA Agreement arts 99 and 100.

[67] “Prime Minister: Britain after Brexit: A Vision of a Global Britain” (2 October 2016), *http://press.conservatives.com/post/151239411635/prime-minister-britain-after-brexit-a-vision-of* [Accessed 16 May 2017].

to the single market. The European Commission makes the decision on whether to grant equivalence. The decision is often based on the technical advice provided by the European Supervisory Authorities (ESA), such as the European Banking Authority (EBA) and the European Securities and Markets Authority (ESMA), though sometimes the Commission itself does all the technical work.[68] Depending on the legislative scheme concerned, the equivalence decision can take the form of an implemented or delegated act.[69] The criteria for recognition of equivalence are set out in the relevant financial legislation. Typically, equivalence provisions require the third country to demonstrate to the regulatory regime that it meets three conditions: (1) it has legally binding requirements in place; (2) it exercises effective supervision; and (3) it achieves the same results as the EU corresponding regime.[70]

The equivalence determination applies legislation by legislation. The following are some of the major activities that are covered by equivalence provisions of EU financial laws:

- providing cross-border investment services to professional clients and eligible counterparties (MiFID II/MiFIR);
- establishing CCPs and information-gathering trade repositories (EMIR);
- providing access to trading venues, CCPs and benchmarks (MiFID II/MiFIR);
- providing access to trading venues for the purposes of trading obligations for derivatives and shares;
- providing marketing of Alternative Investment Funds (AIFMD);
- providing reinsurance (Solvency II).[71]

It must be noted that no equivalence decision can be taken yet under the MiFIR and MiFID II as these two legislations will not come into force until January 2018.[72]MiFID, which is currently the applicable legislation, does not provide for any common third-country regime.[73] Another important issue that can be discerned from the above list is that the equivalence regime does not cover the full range of financial services—deposit-taking, lending, primary insurance, retail asset management and payment services are excluded.[74] Further, even when equivalence is allowed under a particular legislative scheme, equivalence can be granted only partially or provisionally. For example, there are three sets of criteria for

[68] EU Commission, "Equivalence with EU Rules and Supervision", *http://ec.europa.eu/finance/general-policy/global/equivalence/index_en.htm#maincontentSec4* [Accessed 16 May 2017].

[69] Delegated acts are defined as non-legislative acts which are intended to supplement or amend certain non-essential elements of a legislative act. They are considered to be quasi-law-making and are scrutinised by European Parliament and the Council. In contrast, implementing acts are used to set out the uniform conditions for implementing a binding legislation. They are more procedural in nature and their oversight is left to committees of experts or civil servants from member states. See TFEU arts 290 and 291; European Parliamentary Research Service Blog, "Delegated and Implementing Acts" (September 2012), *https://epthinktank.eu/2012/09/24/delegated-and-implementing-acts/* [Accessed 16 May 2017].

[70] EU Commission, "Recognition of Non-EU Financial Frameworks (Equivalence Decisions)", *http://ec.europa.eu/finance/general-policy/global/equivalence/index_en.htm#maincontentSec4* [Accessed 16 May 2017]; House of Lords European Union Committee, "Brexit: Financial services" (December 2016), p.17.

[71] EU Commission, "Equivalence Decisions Taken by the European Commission" (December 2016), *https://ec.europa.eu/info/sites/info/files/file_import/equivalence-table_en.pdf* [Accessed 16 May 2017].

[72] EU Commission, "Equivalence Decisions Taken by the European Commission" (December 2016), *https://ec.europa.eu/info/sites/info/files/file_import/equivalence-table_en.pdf* [Accessed 16 May 2017].

[73] Ernst & Young (EY), "UK/EU: Working through Uncertainty: Practical Considerations for Financial Institutions" (2016), p.8.

[74] House of Lords European Union Committee, "Brexit: Financial services" (December 2016), pp.18, 21.

equivalence under Solvency II: capital requirements, group supervision and reinsurance.[75] A third country's regulatory regime will only be reconciled as fully equivalent when all three criteria are met. To date, only Switzerland and Bermuda have achieved full equivalence, with other third countries—such as Canada, Australia, and Japan—achieving only provisional or partial equivalence.[76] Finally, even when granted, equivalence can be withdrawn on short notice.[77]

The equivalence regime's limited scope and legislation-specific nature makes it a far less attractive option than an EEA-type passport. If the UK decides to pursue equivalence, it must apply to the Commission under the relevant legislative scheme. Since the UK has been an EU member so far and plans to keep a significant portion of the EU legislation through the Great Repeal Bill,[78] achieving equivalence may not be technically difficult at the point of withdrawal. Politically, however, the process for achieving equivalence can prove slow and problematic. An interesting example in this respect is the Commission's landmark decision on equivalence of US central clearing arrangements.[79] Simon Gleeson from Clifford Chance notes that while technical experts found the US regime broadly equivalent within six months, it took more than 30 months of discussion at the political level to grant equivalence. Yet some have remained relatively optimistic that politics will not interfere with equivalence decisions. For example, the Cambridge law scholar Ellis Ferran draws attention to the rise of the ESA and the strong technical expertise that they bring to the equivalence process.[80] In her view, such technical input can shield against political interference. While there is merit to the observation that the ESA play an important role in technical determination, it is only the Commission that can make the equivalence determination. It is hard to see why the Commission's decision cannot be influenced by political considerations, especially when the EU members strongly stress the importance of safeguarding political unity and that the UK should not achieve a better deal than it currently enjoys as a member of the bloc.[81] The UK's withdrawal from the EU, coupled with its reluctance to accept

[75] Directive 2009/138 on the taking-up and pursuit of the business of Insurance and Reinsurance [2009] OJ L335 arts 227, 260, and 172 (Solvency II).

[76] Commission Delegated Decision 2015/1602 of 5 June 2015 on the equivalence of the solvency and prudential regime for insurance and reinsurance undertakings in force in Switzerland based on Articles 172(2), 227(4) and 260(3) of Directive 2009/138, *http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32015D1602&from=EN*; Commission Delegated Decision 2015/2290 on the provisional equivalence of the solvency regimes in force in Australia, Bermuda, Brazil, Canada, Mexico and the United States and applicable to insurance and reinsurance undertakings with head offices in those countries, *http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32015D2290&from=EN*; Commission Delegated Decision 2016/310 on the equivalence of the solvency regime for insurance and reinsurance undertakings in force in Japan to the regime laid down in Directive 2009/138, *http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016D0310&from=EN* [All accessed 16 May 2017].

[77] European Parliament, "Third-country equivalence in EU banking legislation", Briefing PE 587.369 (December 2016), p.2, *http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/587369/IPOL_BRI(2016)587369_EN.pdf* [Accessed 16 May 2017].

[78] See May, "Britain after Brexit" (2 October 2016), *http://press.conservatives.com/post/151239411635/prime-minister-britain-after-brexit-a-vision-of*; Brexit White Paper (February 2017), p.9, *https://www.gov.uk/government/publications/the-united-kingdoms-exit-from-and-new-partnership-with-the-european-union-white-paper* [Both accessed 16 May 2017].

[79] See EU Commission, Press Release, "European Commission Adopts Equivalence Decision for CCPs in USA" (15 March 2016), *http://europa.eu/rapid/press-release_IP-16-807_en.htm* [Accessed 16 May 2017].

[80] Select Committee on the European Union, Financial Affairs Sub-Committee, "Corrected oral evidence: Brexit: Financial Services" (7 September 2016), Q5, *http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/eu-financial-affairs-subcommittee/brexit-financial-services/oral/37866.html* [Accessed 16 May 2017]; E. Ferran, "The UK as a Third Country Actor in EU Financial Services Regulation", *Cambridge Legal Studies Research Paper Series, Paper No.47/2016* (September 2016), p.20.

[81] T. Barber, "The EU 27's Message to Brexit Britain", *Financial Times*, 12 January 2017.

the basic freedoms in return for market access, can adversely affect the prospect of achieving a favourable equivalence deal.[82]

Even if the UK can achieve equivalence in the short term, maintaining such equivalence can be challenging in the long run. It is yet to be seen whether the UK chooses to continue applying EU rules as a third country without any direct influence over the EU regulatory design. If the country decides to break away from EU regulations, it will then find it difficult to maintain the level of equivalence required by EU law. To be sure, international standards continue to apply in both the UK and the EU, providing some level playing-field across jurisdictions. It is unlikely that Brexit will affect the UK's or the EU's commitment to implement standards that have been agreed upon in international fora, such as the Basel Committee on Banking Supervision (BCBS) or the Financial Stability Board (FSB). International standards, however, only lay out minimum requirements that countries can agree upon. Hence, considerable differences remain between jurisdictions in how they regulate and supervise financial markets. As Niamh Moloney rightly notes, equivalence is not simply a matter of regulatory equivalence but also how the authorities supervise firms and enforce the rules.[83] Equivalence assessments for supervisions and enforcement are highly elusive. For example, a tougher or more lenient approach to enforcement on either side can easily diminish the prospect of equivalence.[84] The divergence between the UK and EU regimes is also likely to grow over time as both jurisdictions respond and adapt to fast-paced changes in their own markets. Thus, unless the UK closely follows the EU's lead on financial regulation, losing equivalence will always be a present danger.

A bespoke free trade agreement

The third alternative to EU membership is a bespoke bilateral agreement. This option, which frequently appears in Brexit news and commentaries, seems most aligned with the UK Government's Brexit priorities. In January 2017, the UK Prime Minister indicated that her government sought to pursue "a new, comprehensive, bold and ambitious free-trade agreement" with elements of single market arrangements in certain areas such as financial services.[85] The EU currently has two important bilateral agreements that are relevant to financial services: the EU-Swiss agreement on insurance, and the Canada-European Union Comprehensive Economic and Trade Agreement (CETA).[86]

The EU-Swiss agreement is quite narrow in scope; it only covers a segment of the insurance business that is direct insurance, other than life insurance. Social insurance and reinsurance are also excluded from the agreement.[87] The agreement allows insurance firms which have been licensed by one contracting party to open

[82] S. Hagemann, "Brexit—Six Months On: The Rest of the EU" (27 December 2016), *Ukandeu*, *http://ukandeu.ac.uk/brexit-six-months-on-the-rest-of-the-eu/* [Accessed 16 May 2017].

[83] N. Moloney, "Financial Services, the EU, and Brexit: an Uncertain Future for the City?" (2016) 17 *German Law Journal* 75, 79.

[84] Moloney, "Financial Services, the EU, and Brexit" (2016) 17 *German Law Journal* 75, 79.

[85] "Theresa May's Blueprint for Brexit: Full Speech Transcript", *Financial Times*, 17 January 2017.

[86] Agreement between the European Economic Community and the Swiss Confederation on Direct Insurance Other than Life Insurance [1991] OJ L205/4 (EU-Swiss Agreement); Canada-European Union Comprehensive Economic and Trade Agreement (CETA) (2017), Government of Canada, *http://www.international.gc.ca/trade-commerce/trade-agreements-accords-commerciaux/agr-acc/ceta-aecg/text-texte/toc-tdm.aspx?lang=eng* [Accessed 16 May 2017].

[87] EU-Swiss Agreement, Annex 2, para.A.

a branch in the territory of another contracting party.[88] The agreement, however, does not allow insurance firms to directly provide cross-border insurance services in the territory of the contracting party.[89] The primary supervisory responsibility rests with the home authority where the firm's head office is registered. The home authority sets the solvency margins and verifies the solvency of the undertaking.[90] The host authority, however, can refuse to authorise a branch to pursue business in its territory if it believes that technical conditions for operation have not been met or there are reputational concerns about the firm.[91] This decision can be appealed to a competent court in the host jurisdiction.[92]

CETA, as it appears from its title, is a much broader agreement than the EU-Swiss deal. The agreement, which took more than seven years of negotiation, has been praised as the most comprehensive trade agreement with the EU. It consists of 30 chapters, with the 13th chapter specifically devoted to financial services.[93]CETA seeks to liberalise trade in financial services based on the four modes of freedoms contained in GATS: cross-border supply, consumption abroad, commercial presence and presence of natural persons. It contains the principles of national treatment and most-favoured nation treatment which prohibit contracting parties from discriminating against each other's businesses or treating them less favourably than a third country's firm. Market access under CETA is, however, quite limited as the contracting parties are under no obligation to permit foreign financial institutions to conduct or solicit business in their territory.[94] So, the cross-border provision or consumption of financial services will be subject to the same rules that each contracting party applies in its jurisdiction.[95] Similarly, with respect to commercial presence, financial institutions should comply with the rules that apply in the host jurisdiction. In fact, art.13.6(3)(a) expressly says that "a Party may impose terms, conditions, and procedures for the authorisation of the establishment and expansion of a commercial presence" if it does not lead to discrimination.[96] As a result, a European bank that seeks to operate in Canada has to comply with the all the requirements imposed by the Canadian regulator, including the rule that foreign branches cannot accept deposits of less than CA$150,000.[97] Similarly, a Canadian bank seeking to operate in the EU has to comply with the EU directive on the taking-up and pursuit of credit institutions' business.[98] In this respect, CETA bears resemblance to GATS, which also provides for a prudential carving-out: contracting parties can adopt all necessary regulatory and prudential measures even though they are incompatible with GATS freedoms.

[88] EU-Swiss Agreement art.11.
[89] EU-Swiss Agreement art.7.1, Annex 2, para.B.3.
[90] EU-Swiss Agreement arts 16 and 17.
[91] EU-Swiss Agreement art.11.4.
[92] EU-Swiss Agreement art.14.2.
[93] Government of Canada, "Trade Negotiations and Agreements" (October 2016), *http://www.canadainternational.gc.ca/eu-ue/policies-politiques/trade_agreements-accords_commerciaux.aspx?lang=eng* [Accessed 16 May 2017].
[94] P. Leblond, "CETA and Financial Services: What To Expect?", *CIGI* Papers No.91 (February 2016), p.12, CIGI*https://www.cigionline.org/sites/default/files/cigi_paper_no.91_web.pdf* [Accessed 16 May 2017].
[95] Leblond, "CETA and Financial Services" (February 2016), p.12, *CIGIhttps://www.cigionline.org/sites/default/files/cigi_paper_no.91_web.pdf* [Accessed 16 May 2017].
[96] CETA art.13.6(3)(a).
[97] Office of Superintendent of Financial Institutions (OSFI), "Guide to Foreign Bank Branching", para.8(a), *http://www.osfi-bsif.gc.ca/Eng/fi-if/app/aag-gad/Pages/fbbguide.aspx* [Accessed 16 May 2017].
[98] Directive 2006/48 relating to the taking up and pursuit of the business of credit institutions (as amended) [2006] OJ L177.

As the above discussion suggests, the existing bespoke agreements as they currently stand do not represent good models for post-Brexit negotiations. The Swiss deal is narrow in scope, covering only a small segment of the insurance business. Its passport rights do not allow firms to directly engage in cross-border insurance business. Moreover, given that the insurance industry is not a major user of passport rights, a Swiss-type agreement would be of limited value to the UK's financial sector. CETA is also a poor policy choice given the significant difference of the UK's and Canada's trade profiles. Trade in services is of much greater significance to the UK than to Canada, whose exports to Europe consist mainly of commodities.[99] In fact, the UK runs a substantial trade deficit in goods which it can only seek to compensate with a surplus in services.[100] In contrast to EU or EEA membership, CETA offers very little liberalisation for trade in services. This is particularly the case with respect to the freedom to provide services and the right to establishment or commercial presence, which are crucial to the UK's financial institutions.

It has been argued that the aim of a bespoke agreement should be securing an EU equivalence and passporting rights for the UK's financial institutions. This view has found support with the House of Lords Report on Brexit and Financial Services, which notes:

> "A main purpose of any bespoke agreement, as far as financial services is concerned, will be to supplement the current equivalence regimes to mitigate any loss of access, and to ensure the continuation of equivalence decisions in order to maintain that access."[101]

However, to what extent Brexit negotiations can secure such an aim is yet to be seen. The European leaders have repeatedly emphasised that they will not allow cherry-picking with respect to the single market. It is unlikely that a bespoke agreement can offer passport rights that are only available to EEA members that have accepted the core freedoms and surrendered to the applications of EU law in return for market access. Similar challenges will arise with respect to equivalence. As explained previously, the EU equivalence regime, which is administered by the EU Commission, is relatively new and offers limited market access. Granting a preferential equivalence status and market access to the UK will require changing EU law on equivalence and will raise questions about the consistency of the EU's approach to third countries. Such objectives will therefore be difficult to pursue, requiring extensive negotiations, and will depend on the EU's negotiating position as well.

[99] In 2013, Canada's net export to EU was nearly $33.2 billion. Precious stones and metals accounted for $10 billion and minerals for $4.6 billion. In contrast, exports in services accounted for just under $14.5 billion for the same period. See Canadian Trade Commissioner Service, "Exporting to the EU: A Guide for Canadian Business" (18 January 2017), para.1.4, *Government of Canada*, *http://tradecommissioner.gc.ca/european-union-europeenne/market-facts-faits-sur-le-marche/0000256.aspx?lang=eng* [Accessed 16 May 2017].

[100] For example, in 2014 UK's trade deficit in goods was 6.7% of GDP while its surplus in services for the same period was 4.7%. See, Office for National Statistics, "UK Balance of Payments, The Pink Book: 2016" (29 July 2016), para.4, "Trade", *https://www.ons.gov.uk/economy/nationalaccounts/balanceofpayments/bulletins/unitedkingdombalanceofpaymentsthepinkbook/2016#trade* [Accessed 17 May 2017].

[101] House of Lords European Union Committee, "Brexit: Financial services" (December 2016), p.33.

The path forward—concluding observations

There remain many unanswered questions about Brexit and how it will change the UK's future relationship with the EU and the world. The UK Government's White Paper on Brexit, which came out only after considerable pressure from the UK Parliament, calls for a new strategic partnership with the EU, "including an ambitious and comprehensive Free Trade Agreement".[102] The paper considers it desirable to maintain the deeply integrated trade and economic relationship between both parties.[103] However, it does not specify how such objectives ought to be achieved. There is also no assessment of the trade-offs involved, particularly with respect to the free movement of people. The paper expressly indicates that the Free Movement Directive will cease to apply following Brexit and that "the migration of EU nationals will be subject to UK law".[104] Yet, maintaining the current level of economic and financial integration on the one hand and taking back full control of immigration on the other can prove irreconcilable objectives. Free movement of people is a founding principle of the EU project, enshrined in the EU treaties and case law.[105] As explained previously, except for Lichtenstein, which is a very small country, none of the European countries with access to the single market has managed to impose restrictions on the flow of EU/EEA citizens. Even Switzerland, which has only partial access to the single market, has accepted the free movement of people.[106] It seems unrealistic to expect EU leaders to give the UK access to the single market without having it to accept at least a mild form of free movement in return.

The EU's position was that no negotiations would start until the formal process of leaving the EU was triggered.[107] The UK Government indeed announced its plan to invoke art.50 of the Treaty of the EU no later than March 2017, [108] and it did so on 29 March. As set out in art.50, the EU treaties will cease to apply to the UK when the withdrawal agreement comes into force, or two years from the day the art.50 notification has been made.[109] There has been optimism on both sides that a deal can be achieved within two years.[110] However, it seems unlikely that a withdrawal agreement and a new free trade treaty can be achieved in two years.

[102] Brexit White Paper (February 2017), p.35, *https://www.gov.uk/government/publications/the-united-kingdoms-exit-from-and-new-partnership-with-the-european-union-white-paper* [Accessed 16 May 2017].

[103] Brexit White Paper (February 2017), p.35, *https://www.gov.uk/government/publications/the-united-kingdoms-exit-from-and-new-partnership-with-the-european-union-white-paper* [Accessed 16 May 2017].

[104] Brexit White Paper (February 2017), p.25, *https://www.gov.uk/government/publications/the-united-kingdoms-exit-from-and-new-partnership-with-the-european-union-white-paper* [Accessed 16 May 2017].

[105] C. Mortera-Martinez and C. Odendahl, "What Free Movement Means to Europe and Why It Matters to Britain" (19 January 2017), p.3, *Centre for European Reform*, *https://www.cer.org.uk/publications/archive/policy-brief/2017/what-free-movement-means-europe-and-why-it-matters-britain* [Accessed 17 May 2017].

[106] Mortera-Martinez and Odendahl, "What Free Movement Means to Europe and Why It Matters to Britain" (19 January 2017), p.6, *Centre for European Reform*, *https://www.cer.org.uk/publications/archive/policy-brief/2017/what-free-movement-means-europe-and-why-it-matters-britain* [Accessed 17 May 2017].

[107] See, e.g., "EU Commission Still Refuses UK Talks before Article 50 Triggered" (3 October 2016), *Guardian*, *https://www.theguardian.com/politics/2016/oct/03/eu-commission-still-refuses-uk-talks-before-article-50-triggered* [Accessed 17 May 2017].

[108] Brexit White Paper (February 2017), p.65, *https://www.gov.uk/government/publications/the-united-kingdoms-exit-from-and-new-partnership-with-the-european-union-white-paper* [Accessed 16 May 2017].

[109] This is unless all 27 EU Member States agree to extend the process: Treaty of Lisbon art.50.

[110] P. Foster, "Brexit Deal Could Be Reached By October 2018, Says Lead EU Negotiator Michel Barnier" (6 December 2016), *Telegraph*, *http://www.telegraph.co.uk/news/2016/12/06/eu-brexit-negotiator-michel-barnier-reiterate-no-cherry-picking/*; Brexit White Paper, (February 2017), p.65, *https://www.gov.uk/government/publications/the-united-kingdoms-exit-from-and-new-partnership-with-the-european-union-white-paper* [Accessed 16 May 2017].

Free trade agreements are notoriously slow.[111]CETA, for example, took seven years to negotiate and has not yet come into force.[112] Furthermore, CETA was relatively simple as it did not include the type of service provisions and non-trade tariff barriers that a big service economy such as the UK's needs to negotiate.[113] Significant time is also needed to agree upon important exit matters, such as the UK's exit bill, the rights of EU citizens in the UK and vice versa, and the future of EU agencies located in the UK.[114] Furthermore, it is possible that the future UK-EU agreement will be classified as a "mixed agreement" under EU law, which would then require the lengthy process of domestic ratification by all EU Member States as well.[115]

Taken together, these factors suggest that the UK risks falling off the cliff—that is, leaving the EU without any agreement at all. A clear transition agreement would then be needed to hedge against this risk. The agreement would ideally allow businesses to maintain access to the single market until a new partnership agreement has been put in place.[116] Yet the prospect of a *soft* transitional agreement seems bleak if the UK Government continues to pursue a *hard* Brexit strategy, stressing its determination to leave the single market.[117] In this regard, Michael Barnier, the EU's lead negotiator, has noted that "the term transitional agreement only makes sense if it prepares the way for a future relationship".[118] This statement implies that the terms of the transitional deal need to be aligned with the objectives and terms of the long-term agreement between the parties. If the parties cannot agree upon access to the single market in their post-Brexit partnership, passport rights risk disappearing altogether after the two-year limit of art.50 has been reached. [119]

If the UK leaves the EU without a deal, its relationship with the EU will then fall back on the WTO rules, or more specifically the schedules on goods and

[111] C. Freund and C. McDaniel, "How Long Does It Take to Conclude a Trade Agreement With the US?" (July 2016), *Peterson Institute for International Economics*, *https://piie.com/blogs/trade-investment-policy-watch/how-long-does-it-take-conclude-trade-agreement-us* [Accessed 17 May 2017].

[112] Government of Canada, "Trade Negotiations and Agreements" (October 2016), *http://www.canadainternational.gc.ca/eu-ue/policies-politiques/trade_agreements-accords_commerciaux.aspx?lang=eng* [Accessed 16 May 2017].

[113] J. Denley and D. Roberts, "Reality Check: Will It Take 10 Years to Do a UK-EU Trade Deal Post Brexit?" (15 December 2016), *Guardian*, *https://www.theguardian.com/politics/2016/dec/15/reality-check-will-it-take-10-years-to-do-a-uk-eu-trade-deal-post-brexit* [Accessed 17 May 2017].

[114] Denley and Roberts, "Reality Check: Will it Take 10 Years to do a UK-EU Trade Deal Post Brexit?" (15 December 2016), *Guardian*, *https://www.theguardian.com/politics/2016/dec/15/reality-check-will-it-take-10-years-to-do-a-uk-eu-trade-deal-post-brexit* [Accessed 17 May 2017].

[115] Trade agreements that contain provisions which fall under Member States' responsibility are often referred to "mixed agreements". In addition to the EU, individual Member States also have to ratify mixed agreements according to their national ratification procedures. See EU Commission, "Trade Negotiations Step By Step" (September 2013), p.6, *http://trade.ec.europa.eu/doclib/docs/2012/june/tradoc_149616.pdf* [Accessed 17 May 2017]. The CJEU is soon expected to rule on the legal classification of the EU Singapore Free Trade Agreement (EU SFTA). In an Opinion published on 21 December 2016, A.G. Sharpston found that the EU SFTA is a mixed agreement which can be only concluded by both the EU and Member States acting jointly. See CJEU, "Advocate General's Opinion in Opinion procedure 2/15", Press Release No.147/1621 (December 2016), *http://curia.europa.eu/jcms/upload/docs/application/pdf/2016-12/cp160147en.pdf* [Accessed 17 May 2017].

[116] TheCityUK, "Brexit and UK-Based Financial and Related Professional Services" (January 2017), paras 6–7, *https://www.thecityuk.com/research/brexit-and-uk-based-financial-and-related-professional-services/* [Accessed 17 May 2017]; House of Commons, Treasury Committee, "Oral Evidence: The UK's Future Economic Relationship with the EU" (10 January 2017), Qs 302–306.

[117] "Theresa May's Blueprint for Brexit: Full Speech Transcript", *Financial Times*, 17 January 2017.

[118] A. Barker and J. Brunsden, "Barnier Urges UK to Be Realistic about Trade Terms for Brexit" (6 December 2016), *Financial Times*, *https://www.ft.com/content/e3644d5e-bba9-11e6-8b45-b8b81dd5d080* [Accessed 17 May 2017].

[119] J. Rankin, "EU's Brexit Negotiator Wants to Stop UK Getting 'Soft Transitional Deal'" (29 November 2016), *Guardian*, *https://www.theguardian.com/world/2016/nov/29/eus-brexit-negotiator-wants-to-stop-uk-getting-soft-transitional-deal-michel-barnier-access-single-market* [Accessed 17 May 2017].

services which set out the rights and obligations of the WTO members.[120] Currently, the UK's commitments are integrated in the EU schedules, which means that when the UK leaves the EU it needs to establish its own separate schedules.[121] Under the WTO rules, new schedules can only be established if other WTO members do not oppose them.[122] The UK Government has indicated its plan to minimise any ground for objection by seeking to replicate the existing trade commitments which it currently shares with the EU in the new schedules.[123] However, even if the UK supposedly succeeds in this mission, little access to the EU markets would follow. The EU Schedule Supplement on Financial Services, which has been submitted under the General Agreement on Trade in Services (GATS), offers little market liberalisation.[124] There is no passporting, and no business activities can be pursued in the EU unless in accordance with the EU laws, as well as any other national laws and requirements that apply in the Member State in question.[125] Thus, dropping back on the WTO rules will be significantly detrimental to the UK's financial sector as they by no means constitute a satisfactory substitute for what is currently available under the single market regime.

It is undoubtedly in both parties' interest to preserve access to the single market in their transitional and long-term agreements. There are currently more than 8,000 EU firms that use passporting to access UK markets.[126] As mentioned previously, continental firms raise half their equity and debt through banks based in the UK. Further, the UK accounts for three-quarters of foreign exchange and derivative activity in the UK.[127] Thus, losing access to the UK's deep capital markets can have severe consequences for the EU economy as well. Wolf-Georg Ringe sees this significance of the UK's financial markets as an important bargaining chip. He argues that, while Brexit will inevitably come, it comes more in form than in substance.[128] That is to say, the UK will formally leave the EU but will in substance

[120] In WTO terms, "These schedules contain the commitments made by individual WTO members allowing specific foreign products or service-providers access to their markets. The schedules are integral parts of the agreements." See WTO, "WTO legal texts—Countries' Schedules of Commitments", *https://www.wto.org/english/docs_e/legal_e/legal_e.htm#schedules* [Accessed 17 May 2017].

[121] GATS, European Communities and Their Member States Schedule of Specific Commitments, GATS/SC/31 (1994).

[122] See General Agreement on Trade in Services, 15 April 1994, 1869 U.N.T.S. 183 (entered into force on 1 January 1995), art.XXI.

[123] J. Braithwaite (UK Ambassador and Permanent Representative to the UN and other international organisations in Geneva), "Ensuring a Smooth Transition in the WTO as We Leave the EU", *Foreign & Commonwealth Office*, *https://blogs.fco.gov.uk/julianbraithwaite/2017/01/23/ensuring-a-smooth-transition-in-the-wto-as-we-leave-the-eu/#comment-78593* [Accessed 17 May 2017].

[124] GATS, European Communities and Their Member States Schedule of Specific Commitments, Supplement 4 on Financial Services, GATS/SC/31/Suppl.4/Rev.1 (1999) (EU Schedule Supplement on Financial Services).

[125] In particular, the EU Schedule Supplement on Financial Services explicitly says that third-country branches

> "receive an authorisation to operate in the territory of a Member State under conditions equivalent to those applied to domestic financial institutions of that Member State, and may be required to satisfy a number of specific prudential requirements such as, in the case of banking and securities, separate capitalisation and other solvency requirements and reporting and publication of accounts requirements or, in the case of insurance, specific guarantee and deposit requirements, a separate capitalisation, and the localization in the Member State concerned of the assets representing the technical reserves and at least one third of the solvency margin"

See *EU Schedule Supplement on Financial Services* GATS/SC/31/Suppl.4/Rev.1 (1999), p.2 n.1.

[126] Bailey, Letter from the Financial Conduct Authority to the Andrew Tyrie, Chairman of the Treasury Committee House of Commons (17 August 2016), p.3, *https://www.parliament.uk/documents/commons-committees/treasury/Correspondence/AJB-to-Andrew-Tyrie-Passporting.PDF* [Accessed 16 May 2017].

[127] J. Treanor, "Mark Carney: European Economies Face Hit if Cut off from City of London" (30 December 2016), *Guardian*, *https://www.theguardian.com/business/2016/nov/30/mark-carney-european-economies-face-hit-if-cut-off-from-city-of-london* [Accessed 17 May 2017].

[128] W-G. Ringe, "The Irrelevance of Brexit for the European Financial Market", Legal Research Paper Series, Paper No.3/2017 (January 2017), p.28, *SSRN*, *https://ssrn.com/abstract=2902715* [Accessed 17 May 2017].

remain closely involved in the EU financial market.[129] In his analysis, economic realities and political constellations—including upcoming elections throughout the EU and the UK—make such an outcome inevitable.[130] Along similar lines, a group of experts at Brugal, a Belgium-based think tank, have called for a new "continental partnership" which allows the UK to participate in selected common market policies and maintain access to the single market.[131] Under this proposal, the UK contributes to the EU budget and will have a say in EU law policy-making, though the ultimate decision-making authority remains with the EU.[132] The proposal also finds it essential that the parties accept the rule of law and enforcement measures that safeguard relevant freedoms in the single market.[133]

These assessments and proposals certainly have merit. They recognise that the economic stakes are so high for both parties that they have no choice but to work together to achieve a workable agreement. Moreover, the bulk of negotiations are to be carried out by technocrats who are less concerned with empty political ambitions than with economic realities and practical solutions. Nevertheless, it must be borne in mind that the EU is as much an economic project as an economic one, and that maintaining political unity is currently of greatest importance to EU leaders. It is unrealistic to assume that the UK can simply pursue a hard version of Brexit but at the same time simply copy and paste all the privileges of the single market membership into a tailor-made agreement. It is indeed possible that the UK Government will moderate its stance regarding leaving the single market, that is, leaving the jurisdiction of the CJEU and the abolition of the free movement of EU citizens. However, if such statements are not just political rhetoric meant to please the Leave campaign but rather true objectives and priorities, then a very rocky road to Brexit seems inevitable.

[129] Ringe, "The Irrelevance of Brexit for the European Financial Market" (January 2017), p.28, *SSRN*, *https://ssrn.com/abstract=2902715* [Accessed 17 May 2017].

[130] Ringe, "The Irrelevance of Brexit for the European Financial Market" (January 2017), p.35, *SSRN*, *https://ssrn.com/abstract=2902715* [Accessed 17 May 2017] ibid at 35.

[131] J. Pisani-Ferry et al., "Europe after Brexit: A Proposal for a Continental Partnership" (25 August 2016), p.6, *http://bruegel.org/2016/08/europe-after-brexit-a-proposal-for-a-continental-partnership/* [Accessed 17 May 2017].

[132] Pisani-Ferry et al., "Europe after Brexit: A Proposal for a Continental Partnership" (25 August 2016), pp.6–7, *http://bruegel.org/2016/08/europe-after-brexit-a-proposal-for-a-continental-partnership/* [Accessed 17 May 2017].

[133] Pisani-Ferry et al., "Europe after Brexit: A Proposal for a Continental Partnership" (25 August 2016), p.7, *http://bruegel.org/2016/08/europe-after-brexit-a-proposal-for-a-continental-partnership/* [Accessed 17 May 2017].

’Twixt Cup and Lip: Liability of Traders under Consumer Contracts for Digital Content Damaged in Transit

Jens Krebs[*]
Swansea University

☞ keywords to be inserted by the indexer

Abstract

This article considers trader liability for damage to digital content in transit. Issues arising from the rule introduced by the Consumer Rights Act 2015 are considered and alternatives addressed. The burden of proof and a right to receive a re-supply, regardless of the point of damage, are also considered.

Introduction

In March 2015, the Consumer Rights Act 2015 (CRA) was passed, and it is intended to simplify and improve consumer legislation in the UK. It contains general unfair terms provisions, not only providing a somewhat revised implementation of the Unfair Terms in Consumer Contracts Directive,[1] but also encompassing consumer elements of the Unfair Contract Terms Act 1977 (UCTA) which are removed from UCTA. Alongside its treatment of unfair terms across consumer contracts in general, the Act also addresses the broad range of issues in relation to several specific types of consumer contracts. It deals with contracts for goods, services and “digital content”, and also contracts that are a mixture of these. It is the first piece of UK legislation to specifically recognise digital content as sui generis, distinct from goods and services, and that is dealt with in Ch.3 of the Act, and it is an aspect of this which is addressed here.

This article is, first, concerned with the trader’s liability for damage to digital content during transit. It will discuss the journey of digital content from the trader to the consumer’s internet service provider (ISP), and on to the consumer’s home network, until it finally reaches the end-user device where it would actually be used by the consumer. The aim is to identify the point at which the trader should be regarded as having (at least largely) fulfilled their obligation and ceases to be liable for subsequently occurring damage to the digital content. We will start with the Consumer Rights Act 2015, which stipulates the approach that is now to be applied, and the issues arising from it. We then look for possible alternative ideas involving an analysis of the well-established concept of “passing of risk” for goods, together with a “control”-based approach largely derived from the Proposal for a

[*] LLB, PhD Candidate..
[1] Directive 93/13 on unfair terms in consumer contracts [1993] OJ L95/29.

Regulation on a Common European Sales Law[2] (CESL), before seeking a suggested way forward. The article then moves on to look at the question of whether, even if damage occurs beyond the point where the trader's obligation to supply has been fulfilled, at whatever point that takes place, there might still be some obligation on the trader to attempt re-supply. Finally, an issue relating to the appropriate allocation of the burden of proof will be addressed.

Supplying digital content

We are used to the separation of passing of property and passing of risk in connection with the sale of goods, but when considering liability which traders face when supplying digital content, a different approach was adopted in the Consumer Rights Act 2015. Section 39 CRA focuses on the determination of when digital content is supplied, and it is this test which operates as the equivalent of "passing of risk" in relation to the sale of goods under the Sale of Goods Act 1979 (SGA). Both determine when traders have fulfilled their contractual obligation to supply products of the requisite standard. For the supply of digital content, this occurs if one of two possibilities under s.39(2) CRA is satisfied. This is the case when digital content (a) reaches the consumer's device, or (b) when it reaches a "trader chosen by the consumer" via whose service the content reaches the consumer's device. At first glance the focus seems to be on the supply to the consumer's device. However, the Explanatory Notes make clear that supply to such a "chosen trader" is, in reality, of much greater importance.[3] "Chosen trader", as envisaged under (b), includes, for example, the consumer's ISP, which provides the consumer with connectivity to the internet. By that definition the focus completely shifts from (a) to (b) because virtually all transactions, except those between a consumer and their chosen ISP, have to pass the ISP to reach the consumer's device. Therefore, in most cases traders have fulfilled their obligation before the product has even reached the consumer's modem. Should, however, a situation fall within the scope of (a), the opposite would apply and the consumer would virtually never have to bear the risk of damage during transmission. The Act only offers these two fairly extreme positions, and the next paragraphs will explore the reasons why neither is entirely satisfactory.

Under a scheme of early fulfilment, i.e. s.39(2)(b) CRA, consumers become responsible for the transmission of the content when it is within the charge of the ISP. The Department for Business, Innovation & Skills (BIS) explained that traders would not have to face liability for events which they cannot control.[4] The reverse conclusion would be that the risk is better placed with the consumer for an event which is, in practical terms, as much beyond their as the trader's control, albeit they have chosen the ISP. In such cases, the consumer may be able to rely on the duty of the ISP under the term implied by s.49 CRA which requires a service contract to be performed with reasonable care and skill. This obligation is mandatory and cannot be excluded by the ISP.[5] However, it is only an obligation

[2] Proposal for a Regulation on a Common European Sales Law (2011/0284 (COD)).
[3] CRA s.39(2)(b); cf. Department for Business, Innovation & Skills, "Consumer Rights Act 2015—Explanatory Notes" para.192.
[4] CRA s.39; BIS, "Consumer Rights Act 2015—Explanatory Notes" para.194.
[5] CRA s.39; BIS, "Consumer Rights Act 2015—Explanatory Notes" para.194.

to take "reasonable care and skill", unlike the trader's strict liability in relation to the quality of the digital content, and the only viable option it gives to the consumer is a claim in damages,[6] whereas they are likely to be far happier with a swift re-supply from the trader, and very unlikely to feel able to take action against the ISP.

In fact, it is suggested that an opposite approach might well result in better outcomes. Completing supply, and with it the trader's liability for damage, at a late stage, as under s.39(2)(a) CRA, would be favourable for consumers. They would only have to bear the risk of damage once the digital contents reached their end-user device, and traders would remain liable for any damage until that point. This would mean that traders remained liable for events which are beyond their control, but traders are generally in a stronger position to carry such burden, and re-supplying the product is a low-cost (or very low-cost) remedy which would solve the consumer's problem in many cases. If the problem is persistent, traders might have to find an alternative means of supply (such as a copy on a physical carrier), or try to put pressure on the ISP to resolve the disruption. Eventually, if the problem does originate from a source which is beyond a trader's control, and there is no practicable alternative means of supply, there would be a point at which continuance of the trader's liability, after trying to resend the digital content, might be viewed as unreasonable, but it is arguably much less so than leaving all the risk on the consumer, once the digital content has reached the consumer's ISP. In any event, a much more discriminating approach is suggested below.

Goods—physical form and ownership

We should consider alternatives to the solution offered under s.39 CRA, and first we shall look at the approaches taken to goods. Rules on the passing of risk regarding the sale of goods have been developed over decades.[7] They used to be provided solely by the Sale of Goods Act 1979 (SGA), but the CRA has introduced new rules for the consumer context. Obviously this article is concerned with the latter context in relation to digital content, but both sets of rules for goods will be compared here with the digital situation. The traditional rule under the SGA provides for risk to pass when "property passes", unless the parties have agreed otherwise. This is a rule generally aligning the passing of risk with ownership moving from the seller to the buyer.[8] In contrast, s.29(2) CRA states that, in relation to consumer contracts, risk passes with physical possession of the goods.[9] There are difficulties with carrying over either of these approaches to the context of digital content.

There are obvious problems in adapting to the context of digital content any rule which focuses on physical possession moving from one party to another. The lack of physical form had caused problems in relation to the classification of digital

[6] See CRA s.54(7)(a).

[7] *Goode on Commercial Law*, 4th edn, edited by Ewan McKendrick (London: LexisNexis UK and Penguin Books, 2010), Ch.9.

[8] L.S. Sealy and R.J.A. Hooley, *Commercial Law: Text, Cases, and Materials*, 4th edn (Oxford: Oxford University Press, 2009), pp.70–71.

[9] CRA s.29(2) reads: "The goods remain at the trader's risk until they come into the physical possession of (a) the consumer … ."

content prior to the implementation of the CRA.[10] The argument was that in order for digital content to be goods it had to be tangible, which, in turn, related to a physical presence.[11] To some extent, this was the reason why the Law Commission saw it as necessary to introduce digital content as a sui generis category,[12] and with it emerged the need for a more suitable method of dealing with liability for damage for digital content in transit.

Of course, arguments can be made for some type of physical presence for digital content, but they are difficult in themselves, and unconvincing as the linchpin for a rule about the passing of risk. There is no physical object which departs from the supplier and arrives with the acquirer. Digital content consists of data that exist as a binary sequence on a carrier in the form of bits and bytes. They make up the smallest parts out of which data consists and occupy physical space on the carrier in the form of magnetic or optical status,[13] but it is not the physical object that was dispatched and transferred. They are merely a series of changes to the physical medium by which it is transferred and on which it eventually arrives. Further, the physical aspect of goods goes far beyond that of digital content. For the latter, it is not needed in the same sense as it is for goods; the physical appearance of digital content is a cryptic string of states and mainly an auxiliary means of making the product persistent; and only the correct interpretation of it bears a fruitful result. The core of it is merely an idea translated into binary form, and this is reflected in the fact that digital products are predominantly covered by other areas of law such as intellectual property law, i.e. copyrights, database rights and patents.[14]

The significance of intellectual property rights in relation to digital content not only reflects their intangible nature and the complications in carrying over ideas of transfer of physical possession, but it also provides the core difficulty with transposing the traditional rules from the Sale of Goods Act 1979 to the context of digital content. For business transactions in goods, s.20(1) SGA states that: "Unless otherwise agreed, the goods remain at the seller's risk until the property in them is transferred to the buyer …". Transfer of "property", in turn, means that "there will be transferred to him a title to an absolute legal interest in the goods

[10] In order to determine which rights buyers of digital products would enjoy, it was asked whether digital products were similar to goods or services. Depending on the outcome, buyers would have been granted rights under the SGA or the SGSA. The reasoning provided was, however, never convincing enough to be determinative.

[11] The requirement, even though never expressly stated under UK law, is nicely illustrated in the case of *St Albans City and District Council v International Computers Ltd* [1997] F.S.R. 251CA (Civ Div), where, both at first instance and on appeal, the judges concluded that digital content would only be covered by the SGA, i.e. be classified as "goods", if delivered on a physical medium. Consequently, any other product would lack the necessary physical component to be considered under the SGA. See also K. Moon, "The Nature of Computer Programs: Tangible? Goods? Personal Property? Intellectual Property?" (2009) 31 E.I.P.R. 396; R.Bradgate, "Consumer Rights in Digital Products", Research Report Prepared for the UK Department for Business, Innovation and Skills (September 2010), *https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/31837/10-1125-consumer-rights-in-digital-products.pdf* [Accessed 17 May 2017]; N. Helberger et al., "Digital Content Contracts for Consumers" (2013) 36 J. Consum. Policy 37.

[12] Bradgate, "Consumer Rights in Digital Products" (September 2010), p.16, *https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/31837/10-1125-consumer-rights-in-digital-products.pdf* [Accessed 17 May 2017]; see also Helberger et al., "Digital Content Contracts for Consumers" (2013) 36 J. Consum. Policy 37, 39.

[13] This is true for any digital data; even if they are not saved on a medium they would, for the duration of their existence, still occupy space on the RAM (random access memory) which, in turn, translates to occupation of a certain number of electronic parts, i.e. transistors.

[14] A. Murray, *Information Technology Law* (Oxford: Oxford University Press, 2010), Ch.8.

sold [(ownership)]”.[15] In transactions in relation to digital content, acquirers, especially consumers, virtually never obtain an absolute legal title in the product.[16] Ownership commonly remains with the producer, and consumers pay for a licence to exploit the contents in question. Such licences are not “transferred”, but rather come into existence at the time of purchase. In some sense, they are the main substance of the contract without which the consumer would not obtain anything usable.[17] Transfer of ownership does not generally take place when the product is digital.

Therefore, neither the approach on the “passing of risk” involving physical possession under the Consumer Rights Act 2015, nor that relating to ownership under the Sale of Goods Act 1979, offers a solution which could simply be adopted to answer the question of the extent of the endurance of the trader’s liability for damage to digital content during transit. There is nevertheless one desirable aspect of the CRA which should be emphasised: the rule is not at the mercy of the normally superior bargaining power of a commercial seller because, apart from a very few exceptions, the rule in s.29 CRA is unaffected by express agreement and intention to the contrary.[18] This aspect is equally true under s.39 CRA, and should certainly not be changed.

The European approach

We should now turn to another alternative approach to dealing with our problem. In 2011 the European Commission put forward an approach in their Proposed CESL.[19]Article 142.2 addressed the “passing of risk” in consumer contracts for digital content and stated:

> “In a contract for the supply of digital content not supplied on a tangible medium, the risk passes at the time when the consumer or a third party designated by the consumer for this purpose has obtained the control of the digital content.”

Just like s.39 CRA, this provision avoids the problematic concepts of “possession” or “ownership”, but it adopts a different approach from that of the CRA. Of course, it raises the question of the meaning of “a third party designated by the consumer”, and if this phrase is understood also to extend to the consumer’s ISP, the approach

[15] *Benjamin’s Sale of Goods*, ?? edition, edited by M.G. Bridge and J.P. Benjamin (London: Sweet & Maxwell/Thomson Reuters, 2012), para.5-003 (check edition [1992]); also Goode, *Goode on Commercial Law* (2010), p.267.

[16] The only exception would be a purchaser who commissioned the development of a particular product. In such a situation it is potentially intended that the developer transfers not only a copy of the content, but the entire source code, a compiled version and all rights attached.

[17] D. Rowland, U. Kohl and A. Charlesworth, *Information Technology Law*, 4th edn (Abingdon: Routledge, 2012), pp.398–401.

[18] In comparison, SGA s.20(1), which is concerned with the passing of risk in business transactions, begins with the words “Unless otherwise agreed”, which are not featured in CRA s.29.

[19] The proposal has to be treated with care as it was withdrawn in its entirety on 16 December 2014 owing to the idea of producing a “[m]odified proposal … to fully unleash the potential of e-commerce in the Digital Single Market” (European Commission, “Commission Work Programme 2015”, Annex 2, Item 60). The Commission has since launched a much larger campaign: European Commission, “Digital Single Market—Bringing down barriers to unlock online opportunities”, *http://ec.europa.eu/priorities/digital-single-market/* [Accessed 18 May 2017]. As part of this broader work on the Digital Single Market, a Proposal for a Directive on certain aspects concerning contracts for the supply of digital content (2015/0287 (COD)) has been brought forward. Nevertheless, the potential of what was proposed in CESL as a model should still be considered.

has the same problems as s.39 CRA.[20] However, this approach would seem to have been intended to require the consumer to actively "designate" another party who is to receive the digital content (i.e. a true agent for such receipt), and we will proceed on this basis. Whether or not it is correct, as an interpretation of the proposed art.142.2 it is an approach which provides a useful contrast. It could circumvent those limitations encountered under s.39 CRA and be advantageous to consumers, as the following will show, but it does depend strongly on the notion of "control". Prima facie, this concept exhibits greater similarities to the pragmatic concept of possession than the more esoteric concept of ownership, and might be seen to reflect the reasonable expectations of consumers. Furthermore, it was designed to be "mandatory in nature; which means that the parties may not, to the detriment of the consumer, exclude its application or derogate from or vary its effects".[21] Thus, it should not be subject to change through the greater bargaining power of the trader.

Blázquez has considered the concept of control, explaining his views by reference to the idea of "full control". In this case, consumers would take "control" of the product once it has reached a location where it would reasonably be used as intended by the consumer. In other words, even though digital content could potentially be manipulated within the consumer's network while in transit, this would locate "control" on the consumer's end-user device. Only when the intact product arrived at such a device would the trader be released from their obligation to supply. Blázquez asserts that art.142.2 CESL aims to achieve this and the correct interpretation of the provision would be that "risk is [only] borne by the consumer when he has obtained full control of the digital content".[22] Of course, "full control", as suggested by Blázquez, would be identical to the position under s.39(2)(a) CRA and suffer from the same weaknesses. Placing liability on the trader for the entire transfer is certainly not ideal but would be better than placing the entire burden on the consumer at a much too early stage; but "control" need not be treated in this way. A much more discriminating approach could be taken to it.

First, consumers could be regarded as in control of the content once it has passed the home router (or modem),[23] which would mean that consumers have to bear the risk of any interference within their internal network, but this should also be seen as too blanket an approach, and heed has to be paid to the possible complications which may arise once the digital content has passed the consumer's modem in order to address the point at which risk should lie at the consumer's door.

Along the way from the modem to the point where the content is accessible, a multitude of events can occur which might compromise the product. Consumers might, for instance, utilise demilitarised zones (DMZ) or proxy servers, technical facilities which perform security tasks within a private network. Much more

[20] The proposed Digital Content Directive mirrors s.39 CRA in its impact (see art.5 and Recital 23). It could be questioned as to whether this is a reaction to the approach taken to "control" by Blázquez (see text at fn.22), under which two extreme positions, but this time favouring the consumer, in relation to the point of liability of the trader would again have been generated. The suggestion being made here is for a much more discriminating approach, more appropriately dividing the burden of damage in transit between trader and consumer.

[21] F. Oliva Blázquez, "Passing of Risk" in J. Plaza Penadés and L. M. Martínez Velencoso (eds), *European Perspectives on the Common European Sales Law* (Switzerland: Springer International, 2014), p.201.

[22] Blázquez, "Passing of Risk" in *European Perspectives on the Common European Sales Law* (2014), p.201 (emphasis added).

[23] Most domestic networks will now use a router with an integrated modem to connect their devices to the internet. Yet the modem must be understood to be the interface between the two spheres.

common are anti-virus scanners and software firewalls which operate on the consumer's laptop, tablet or phone and, in some cases, are even built into the operating system.[24] The risks in relation to transmission between modem and end-user device are too diverse for a blanket solution thereafter and we need to look at what this should turn upon.

We should consider an event-based approach, where "control" over digital content depends on the component of the consumer's system which causes the damage. It is contended that we should look at what can be technologically expected of the "average consumer" to distinguish these events, and determine where risk should lie in relation to them. The average consumer will normally employ certain components for the reasonable protection of his or her private network, such as firewalls and anti-virus scanners[25]; other components are implemented as a more particular choice of the individual consumer, in the form of more advanced technology to enhance the security of their private network, the use of which exceeds the knowledge and abilities of the average consumer. If not carefully set up, such means carry an increased chance of damage to otherwise flawless content. The components typically falling within this group are DMZs and proxy servers. Generally, consumers who decide to enhance their network protection beyond the average consumer's standard are aware of the technical risks involved and proceed in contemplation of them. Thus, a distinction could be drawn, making traders not liable for damage arising from such components, but relieving consumers of bearing the risk of interfering events which are caused by components that the average consumer would employ. The view can be taken that traders are aware of the use of anti-virus scanners and firewalls, and their products should be developed to a standard that would not cause problems with the security components used, knowingly or otherwise, by the average consumer. Of course, not only the type of component, for example an anti-virus scanner, but also its quality must be such that the average consumer would employ it, depending upon such factors as price, producer's reputation and, naturally, market share, if the trader is to be liable.

Obligation to re-supply

We have looked at the point at which the risk of damage to the digital content in transit should be transferred from the trader to the consumer. Now it has to be considered whether, at whatever point that occurs, an obligation to re-supply a digital product which was damaged in previous attempts of supply should continue for a time after that has occurred.

Satisfactory quality

First, we should briefly acknowledge an argument which might be made in relation to using the implied term that digital content should be of "satisfactory quality".

[24] See Microsoft Corp, "Windows Firewall", *http://windows.microsoft.com/en-gb/windows-8/windows-firewall-from-start-to-finish*; also Microsoft Corp, "Windows Defender", *http://windows.microsoft.com/en-gb/windows/using-defender* [Both accessed 18 May 2017].

[25] Software firewalls and anti-virus applications can be deactivated temporarily or even removed, but it would be unreasonable to set the threshold at such a high level, where consumers need to have in-depth knowledge, especially when it is recommended for the protection of the consumer's data to keep such means active at all times. Furthermore, the supply of digital content can occur in so many different forms (via email, downloads, web access, cloud access, etc.) that the recurring need to deactivate the protective measures would compromise the system's security.

Section 34 CRA implies a term into consumer contracts for the supply of digital content which requires digital content to be of satisfactory quality. The test under s.34 CRA is derived from the well-known test for satisfactory quality for sales of goods under s.14 SGA. It sets out an exercise requiring consideration of the "standard of a reasonable person" against the background of the price, description and any other relevant factors. The applied standard should encompass, inter alia, freedom from security holes as well as freedom from technical flaws and defects,[26] but what should be addressed here in particular is the durability aspect of satisfactory quality. Part of the quality assessment requires digital products, in appropriate cases, to be durable.[27] In relation to goods, in the case of *Lambert v Lewis*,[28] Lord Diplock examined the implied term of what was then "merchantability" in relation to durability, and came to the conclusion that duties as to "merchantable quality", as imposed by the Sale of Goods Act 1979, would continue "for a reasonable time after delivery".[29] An argument might be encountered that "durability" imposed a duty for traders to supply a product which would continue to be of satisfactory quality for a reasonable time. It would then cover those incidents affecting digital content after the point when it has been supplied. However, such an argument should be quickly dismissed. Durability is only concerned with a problem which was part of the product before the supply, but which simply did not manifest itself until after that point.[30]

Licences

Consideration should thus be given to a more holistic possibility for an obligation to re-supply, also covering cases where a problem occurred after the "supply" of the digital content, but before it reached the device on which the consumer would use it. The significance of licences for the supply of digital content must be addressed.

As was explained before, consumers usually do not purchase the digital content (or the copyright in it) but a licence that permits its use.[31] In online transactions, the licence agreement is commonly presented to the consumer during the contracting process, as a so-called click-wrap licence.[32] If the digital content is damaged in transit, then prima facie the consumer has a licence but nothing usable under it. The existence of such a licence gives rise to an argument for an obligation to re-supply.

What needs to be looked at is whether, in such circumstances, traders impliedly agreed to re-supply (or make available) the contents after the purchase. In a somewhat different context this type of idea was considered by Jacob J in *Sony*

[26] N. Helberger et al., *Digital Consumers and the Law* (Alphen aan den Rijn: Kluwer Law International, 2013), p.106.

[27] CRA s.34(3)(d).

[28] *Lambert v Lewis* [1982] A.C. 225 HL. Lord Diplock's analysis of this case is of particular interest as the agreement was between a trader and a consumer.

[29] *Lambert v Lewis* [1982] A.C. 225 at 276; see also *Benjamin's Sale of Goods*, para.11-057.

[30] Cf. *Mash & Murrell Ltd v Joseph I Emanuel Ltd* (1961) 1 W.L.R. 862 CA.

[31] S. Stokes, *Digital Copyright*, 4th edn (Oxford: Hart Publishing Ltd, 2014), p.140.

[32] It is commonly referred to as a "click wrap licence". However, this is not the right place to embark upon the exact meaning to be accorded to such terminology.

Computer Entertainment Inc v Owen,[33] a case involving digital content provided on a physical carrier. In his judgment, Jacob J explained that

> "if you got your licence and your [product] would not work, you would still have a licence, but you could not operate your licence".[34]

The tenor of his statement suggests that there ought to be something in the contract allowing the consumer to make use of this licence. This could be seen as an appropriate situation for an implied term, and one implied in law, as suitable to all contracts of a given type, rather than in fact, i.e. just on the basis of the intention of those particular parties. Implying terms in law requires identification of a type of contract, and here it would seem to be consumer contracts for the supply of digital content. The other requirement for such an implication has been stated as one of "necessity".[35] However, as much analysis has shown,[36] and the courts are beginning to recognise,[37] this would be better stated simply as a matter of "reasonableness, fairness and the balancing of competing policy considerations".[38] It then becomes a question of whether the consumer should be left with a useless licence, or whether there should be some obligation on the trader to make some attempts to re-supply on the consumer's request.

Of course, in arguing for such an implied term, it should be asked whether an obligation to re-supply would place an unreasonable burden on traders. There are two factors to be taken into account, which are "effort" and "cost". When considering "effort", one must look at the period during which the obligation would persist and the number of attempts the trader might have to make. Traders might think it unreasonable for such an obligation to persist over the entire period during which the licence is valid, but consumers who cannot make use of their licence because they never received a working product might expect there to be such an obligation. On the other hand, if, after numerous attempts, the consumer cannot resolve the problem with their system in order to allow a successful supply, a trader might argue they should be regarded as having fulfilled their contractual obligation. It becomes a question of balancing the amount of effort a trader reasonably has to put into attempts to re-supply, against the expectations of the consumer. In many instances, human intervention would not be required at all on the part of the trader. If the digital content is accessible online by the consumer, it could eliminate the need for action by the trader completely. Consumers could request the product from a server numerous times until it eventually reached the end-user device unharmed. This constant availability of digital content obviously also helps in relation to the costs for this "service". Unlike goods, digital content can simply be sent, or downloaded, at a negligible cost.

Under an implied term to re-supply, the obligation should be maintained for a reasonable time, but as it can be a low "effort" and "cost" obligation, the time

[33] *Sony Computer Entertainment Inc v Owen* [2002] E.M.L.R. 742 Ch D.
[34] *Sony Computer Entertainment v Owen* [2002] E.M.L.R. 742 at 748.
[35] Per Lord Wilberforce in *Liverpool City Council v Irwin* (1976) 2 All E.R. 39 HL at 44.
[36] E. Peden, "Policy Concerns behind Implication of Terms in Law" (2001) 117 L.Q.R. 459, 466–467.
[37] *Crossley v Faithful & Gould Holdings Ltd* (2004) 4 All E.R. 447 CA (Civ Div) at [36].
[38] *Treitel on the Law of Contract*, 14th edn, edited by E. Peel (London: Sweet & Maxwell, 2015), para 6-035; *Koffman and Macdonald's Law of Contract*, 8th edn, edited by E. Macdonald and R. Aktins (Oxford: Oxford University Press, 2014), para.7.43; *Atiyah's Introduction to the Law of Contract*, 6th edn, edited by P.S. Atiyah and S. A. Smith (Oxford: Oxford University Press, 2005), p.161.

frame should be applied generously to enable, at least, the initial operation of the consumer's product licence. Imposing such an obligation, even extensively, does not appear to burden traders in a significant manner, but it might raise concerns of susceptibility to abuse by consumers. However, technological limitations can be imposed to prevent consumers who received their original supply without problems from taking advantage of re-supply to try to obtain a second useable copy. We are familiar with the use of digital rights management (DRM) allowing traders to control the use of digital content by limiting or monitoring access, including attempts to duplicate,[39] and the same could be used here. Moreover, products are often "unlocked" or "redeemed" on a user account for an online service via which digital content can be accessed for the licence duration. Should consumers then try to acquire a second copy, DRM would restrict access to prevent simultaneous or unauthorised use. The potential for abuse is not a significant argument against the law imposing an obligation to re-supply digital content on traders.

Burden of proof

One final issue regarding the liability of traders, which is considered here, must be addressed and that is in relation to the burden of proof. As is generally the case, especially for the ordinary person, it is difficult to provide evidence of a causal link between technological mishaps and the damage in question. Being unable to establish this link is likely to leave consumers without the possibility of enforcing their rights when traders deny responsibility. In the light of this, the burden of proof should be placed on traders who are, in any event, likely to have superior technological expertise or resources.[40] If damage occurs, the trader should have to show that either the faulty component is one which deviates from the standard of the average consumer, and hence is regarded as within the consumer's "control", or, at least, that none of the other components in relation to which the trader would be liable, if those components were responsible, has caused the damage. In other words, effectively the trader's liability is presumed, but can be rebutted by identifying the faulty component as one for which the consumer is responsible, such as one not used by the average consumer, or eliminating the components for which the trader is liable, such as one used by the average consumer. The recently proposed Digital Content Directive[41] suggests this approach with regard to the conformity of digital content,[42] which also relates to failure to supply.[43] Of course, there is an issue in relation to the trader having access to the components which are within the consumer's system. The consumer may be unwilling to co-operate with an investigation by the trader or, as is more likely, unable to provide the assistance which the trader needs to do so. Such a scenario is also considered by the proposed Directive.[44] It provides for the burden not to fall on the trader where,

[39] C. May, *Digital Rights Management: The Problem of Expanding Ownership Rights* (Oxford: Chandos Publishing (Oxford) Ltd, 2007), pp.67–69.

[40] Proposal for a Directive on certain aspects concerning contracts for the supply of digital content (2015/0287 (COD) (pDCD), Recital 32.

[41] pDCD. See further fnn.19 and 20 above.

[42] pDCD art.9.

[43] pDCD art.10(1).

[44] pDCD art.10(3) and Recital 32.

after having made "necessary efforts", the trader cannot obtain the required access to components. The proposed Directive provides guidance as to the extent to which traders are required to act. "Necessary efforts" can be understood to require some form of positive action which must be proportionate to the situation, the product and the method of delivery.[45] As such, it would be reasonable to contact the consumer in order to obtain information about the set-up of the network and its components, but, to safeguard the consumer's privacy, a trader's first response should not be direct remote access to the consumer's device, and Recital 33 states that the consumer "should cooperate … in order to allow the [trader] to ascertain the consumer's digital environment". Having made those efforts, if the trader shows that all the components which are accessible to the trader, for which they are liable, have not caused the damage, they will not be responsible, unless the consumer provides evidence to the contrary.

Conclusion

The problem of damage to digital content in transit has been considered, and a number of ways of dealing with it have been addressed. It is contended that a preferable solution to that in the current legislation can be found. Section 39 of the Consumer Rights Act 2015 addresses the trader's liability for damage in relation to digital content during transfer, but the method it adopts lacks sufficient discrimination and favours traders in far too many situations.

However, the understanding of the proposed CESL discussed here, under which "risk passes" when buyers gain control over the digital content, provides an apt starting point which only needs to be refined in terms of the meaning of "control", and this can take place by reference to what can be expected of the average consumer. Alongside that, the suggested approach to the burden of proof should also be adopted. However, at whatever point in the transfer "risk passes", the situation would be assuaged by an obligation to re-supply digital content which has not reached the point of being usable on the consumer's end-user device undamaged.

[45] pDCD, Recital 33.

Towards a Consistent Use of Subsequent Conduct in Singapore Contract Law

Goh Yihan[*]

Singapore Management University

keywords to be inserted by the indexer

Abstract

This article, through the example of Singapore law, questions whether the various uses of subsequent conduct in contract law are consistent with each other. It will be shown that there is in fact a great deal of inconsistency and that any debate about the proper ambit of subsequent conduct within each specific use must necessarily yield to this prior question of consistency. In order to advance this main point, this article will be organised in the following way. It will first discuss the uses of subsequent conduct in Singapore contract law and explain why they are inconsistent with one another. Following from this, this article will argue that there are cogent grounds founded on principle that require a consistent use of subsequent conduct across contract law. However, the requirement of consistency does not tell us what the appropriate and consistent approach ought to be. The final part of this article will therefore argue for an expansive use of subsequent conduct in contract law, as aptly justified by policy grounds.

Introduction

Contracting parties' conduct subsequent to the formation of their contract, or their "subsequent conduct", has several potential uses in contract law. For one, it may be used to ascertain the existence and terms of the original agreement that was made orally or only partially expressed.[1] Flowing from that, it can certainly be used to infer a separate agreement that may have varied the original contract or given rise to an estoppel. Finally, it may also be used to interpret the meaning of the original contract. It is of course well established that English law has ruled out this last possible use of subsequent conduct,[2] but other common law jurisdictions have not quite done so.[3] While these uses of subsequent conduct in contract law are internally well settled, a more fundamental question needs to be asked regarding

[*] Goh Yihan, School of Law.

[1] *Ferguson v Dawson & Partners (Contractors) Ltd* [1976] 1 W.L.R. 1213 CA (Civ Div); *Wilson v Maynard Shipbuilding Consultants AB* [1978] Q.B. 665 CA (Civ Div); *Mears v Safecar Security Ltd* [1983] Q.B. 54 CA (Civ Div) at 77; *Carimichael v National Power Plc* [1999] 1 W.L.R. 2042 HL at 2050–2051; and *Thorner v Major* [2009] 1 W.L.R. 776 HL.

[2] *James Miller & Partners Ltd v Whitworth Street Estates (Manchester) Ltd* [1970] A.C. 572 HL at 603, 606; *L Schuler AG v Wickman Machine Tool Sales Ltd* [1974] A.C. 235 HL.

[3] See, e.g., *Gibbons Holdings Ltd v Wholesale Distributors Ltd* [2008] 1 N.Z.L.R. 277.

the consistency across these uses. For example, is the steadfast rejection of subsequent conduct for contractual interpretation necessarily consistent with the ready use of such evidence to infer the existence of a contract?

This article, through the example of Singapore contract law—which is influenced by both English contract law and aspects of Indian (statutory) evidence law—questions whether the various uses of subsequent conduct are consistent with each other. It will be shown that there is in fact a great deal of inconsistency and so any debate about the correctness of each specific use of subsequent conduct must necessarily yield to this prior question of consistency. In order to advance this main point, this article will be organised in the following way. It will first discuss the uses of subsequent conduct in Singapore contract law and explain why some uses are inconsistent with one another. This article will then argue that there are cogent grounds founded on principle that require a consistent use of subsequent conduct across contract law. However, the requirement of consistency does not tell us what the *appropriate* approach ought to be. The final part of this article will therefore argue for an *expansive*—in the sense that there should not be any legal restrictions—use of subsequent conduct in contract law, as aptly justified by both principle and policy grounds.

The inconsistent uses of subsequent conduct in Singapore contract law

Contractual interpretation

Previous approach—prohibition of subsequent conduct for interpretation

The use of subsequent conduct for the interpretation of a contract in Singapore law has undergone a fundamental change of late. Originally, Singapore law, following English law, disallowed the use of subsequent conduct for the interpretation of contracts. Thus, the Singapore High Court[4] held in *Re Lin Securities (Pte) Ltd*[5] that it was "not entitled to rely upon anything which any party subsequently did or said" in construing the contract in question. In doing so, the court cited Lord Reid's well-known words in *James Miller & Partners v Whitworth Street Estates (Manchester) Ltd*[6]:

> "I must say that I had thought that it is now well settled that it is not legitimate to use as an aid in the construction of the contract anything which the parties said or did after it was made. Otherwise one might have the result that a contract meant one thing the day it was signed, but by reason of subsequent events meant something different a month or a year later."

In addition, the court also referred approvingly to the subsequent House of Lords decision of *L Schuler AG v Wickman Machine Tool Sales Ltd*,[7] which not only affirmed *James Miller*, but also became one of the leading authorities in English

[4] Hereafter referred to as the High Court unless the context requires otherwise.
[5] *Re Lin Securities (Pte) Ltd* [1988] 1 S.L.R.(R.) 220.
[6] *James Miller & Partners v Whitworth Street Estates* [1970] A.C. 583 at 603 (*James Miller*).
[7] *L Schuler v Wickman Machine Tool Sales* [1974] A.C. 235 (*Schuler*).

law for the proposition that the contracting parties' subsequent conduct cannot be used for the interpretation of their contract. After *Re Lin Securities (Pte) Ltd*, both the High Court[8] and the Singapore Court of Appeal[9] (the highest court in the land) have endorsed the application of this proposition in Singapore on multiple occasions. Indeed, by 2006, the High Court was able to say in *Panwah Steel Pte Ltd v Koh Brothers Building & Civil Engineering Contractor (Pte) Ltd* that the proposition in *James Miller* "has been applied in Singapore in many cases".[10]

The Singapore courts have advanced various policy reasons to support this proposition. One key reason given is that the meaning of the contract is fixed at the point of formation, and so the parties' conduct after that point cannot logically affect the meaning already reached. In *Jia Min Building Construction Pte Ltd v Ann Lee Pte Ltd*, the High Court illustrated this idea by describing reliance on subsequent conduct in interpreting a contract as being akin to "having the tail wag the dog".[11] Another reason given by the courts is that the use of subsequent conduct will result in uncertainty, as the parties can never be sure of their original bargain. This is evident from the Singapore courts' repeated citation of Lord Reid's speech in *James Miller*, where his Lordship spoke disparagingly of the possibility that the meaning of a contract could be changed by reason of subsequent events.[12] A related reason advanced is that consideration of the parties' subsequent conduct is akin to considering their subjective intentions, which is not permitted under the objective theory of contractual interpretation. Thus, the High Court in *Re EG Tan & Co (Pte)* appeared to equate the parties' subsequent conduct with their "subjective wishes or intent of either party", which the court rightly held was eschewed in the interpretative process.[13] The same court put this reason across more strongly in *Standard Chartered Bank v Neocorp International Ltd* when it said that subsequent conduct evidenced the parties' subjective intentions and was "not only almost invariably wholly unreliable but often also contain an element of posturing", making it incompatible with the objective approach.[14]

We shall have occasion to revisit these policy reasons later, but for now the point is that Singapore law had, until barely 10 years ago, steadfastly followed the English position that subsequent conduct could not be considered in interpreting a contract. This settled approach was to take a fundamental change following the Court of Appeal's seminal decision of *Zurich Insurance (Singapore) Pte Ltd v B-Gold Interior Design & Construction Pte Ltd*[15] in 2008.

[8] See, e.g., *Las Vegas Hilton Corp (t/a Las Vegas Hilton) v Khoo Teng Hock Sunny* [1996] 2 S.L.R.(R.) 589 at [49]; *Kredietbank NV v Sinotani Pacific Pte Ltd (Agricultural Bank of China, third party)* [1999] 1 S.L.R.(R.) 274 at [113]; *Encyclocom Education Pty Ltd v Horizoneducom Pte Ltd* [2003] 4 S.L.R.(R.) 165 at [21]; *Rickshaw Investments Ltd v Nicolai Baron von Uexkull* [2006] 2 S.L.R.(R.) 850 at [15]; and *Chia Kok Kee v HX Investment Pte Ltd* [2007] SGHC 164 at [52].

[9] Hereafter referred to as the Court of Appeal unless the context requires otherwise. See, e.g., *National Employers' Mutual General Insurance Association Ltd v Globe Trawlers Pte Ltd* [1991] 1 S.L.R.(R.) 550 at [23]; *Estate of Seow Khoon Seng, deceased v Pacific Century Regional Developments Ltd* [1997] 1 S.L.R.(R.) 1 at [27]; *MAE Engineering Ltd v Fire-Stop Marketing Services Pte Ltd* [2005] 1 S.L.R.(R.) 379 at [41]; and *Singapore Telecommunications Ltd v Starhub Cable Vision Ltd* [2006] 2 S.L.R.(R.) 195 at [26].

[10] *Panwah Steel Pte Ltd v Koh Brothers Building & Civil Engineering Contractor (Pte) Ltd* [2006] 1 S.L.R.(R.) 788 at [8].

[11] *Jia Min Building Construction Pte Ltd v Ann Lee Pte Ltd* [2004] 3 S.L.R.(R.) 288 at [41].

[12] *James Miller* [1970] A.C. 583 at 603.

[13] *Re EG Tan & Co (Pte)* [1990] 2 S.L.R.(R.) 386 at [14]. See also *Jia Min Building Construction v Ann Lee* [2004] 3 S.L.R.(R.) 288 at [41].

[14] *Standard Chartered Bank v Neocorp International Ltd* [2005] 2 S.L.R.(R.) 345 at [44].

[15] *Zurich Insurance (Singapore) Pte Ltd v B-Gold Interior Design & Construction Pte Ltd* [2008] 3 S.L.R.(R.) 1029 (*Zurich Insurance*).

Current approach—unsettled use of subsequent conduct for interpretation

It is no exaggeration to say that *Zurich Insurance* has fundamentally changed the way contracts are interpreted under Singapore law. While the Singapore courts had already endorsed Lord Hoffmann's influential attempt to liberalise contractual interpretation in *Investors Compensation Scheme Ltd v West Bromwich Building Society*,[16] *Zurich Insurance* took that endorsement to the next level. Writing for a unanimous court, V.K. Rajah JA held that the admissibility of extrinsic evidence to interpret contracts in Singapore is a "pragmatic and principled one".[17] His Honour also held that the Singapore courts adopt "the modern contextual approach to interpretation, in line with the developments in England in this area of the law to date".[18] Under this approach as applied in Singapore, extrinsic evidence is admissible so long as "it is relevant, reasonably available to all the contracting parties and relates to a clear and obvious context".[19] These requirements have since become known as the "*Zurich Insurance* requirements" and are necessary conditions to fulfil before extrinsic evidence may be admitted to interpret contracts. More pertinent for present purposes is Rajah JA's statement that "there should be no absolute or rigid prohibition against evidence of previous negotiations or subsequent conduct", although his Honour also noted that

> "the relevance of subsequent condict remains a controversial and evolving topic that will require more extensive scrutiny by [the Court of Appeal] at a more appropriate juncture".[20]

Zurich Insurance therefore not only affirmed—indeed, advanced—the contextual approach in Singapore, it also opened up the possibility that subsequent conduct could, in appropriate circumstances, be admitted for the purpose of interpreting contracts under Singapore law.

The problem though is that *Zurich Insurance* did not settle once and for all whether and when subsequent conduct can be used to interpret contracts. To be fair, this was not an issue directly before the Court of Appeal in *Zurich Insurance* and so it rightly declined to deal with it more determinatively. The outcome, however, is that the lower courts are still divided on the relevance of subsequent conduct up to the present day. The first view adopted by the High Court is that subsequent conduct *cannot* be used at all. This is typified by the High Court decision of *Lian Hwee Choo Phebe v Maxz Universal Development Group Pte Ltd*, where the court did not even discuss *Zurich Insurance* and simply applied *James Miller* for the proposition that subsequent conduct cannot be admitted to interpret contracts under all circumstances.[21]

The second view taken by the High Court is that subsequent conduct can *always* be used, either without any or with only limited restrictions. Such a view can be seen in *Goh Guan Chong v AspenTech Inc*, where the High Court thought that *Zurich Insurance* did not lay down any "absolute or rigid prohibition" against the

[16] *Investors Compensation Scheme Ltd v West Bromwich Building Society* [1998] 1 W.L.R. 896 HL.
[17] *Zurich Insurance* [2008] 3 S.L.R.(R.) 1029 at [132].
[18] *Zurich Insurance* [2008] 3 S.L.R.(R.) 1029 at [132].
[19] *Zurich Insurance* [2008] 3 S.L.R.(R.) 1029 at [132].
[20] *Zurich Insurance* [2008] 3 S.L.R.(R.) 1029 at [132].
[21] *Lian Hwee Choo Phebe v Maxz Universal Development Group Pte Ltd* [2008] 4 S.L.R.(R.) 265 at [16]–[17].

admissibility of prior negotiations and subsequent conduct.[22] While this is a literal application of *Zurich Insurance*, it is significant that the court held that previous contractual drafts, which it regarded as a type of prior negotiations, can be admitted even if they will not always be helpful.[23] Given that *Zurich Insurance* had analysed both prior negotiations and subsequent conduct together, the court's expansive allowance of prior negotiations must also apply to subsequent conduct. In *Fico Sports Inc Pte Ltd v Thong Hup Gardens Pte Ltd*, the High Court stated with similar certainty that the "extrinsic evidence that is admissible [for contractual interpretation] includes prior negotiations and subsequent conduct", without any qualification.[24] The High Court has imposed some limited restrictions in other cases. For example, in *Leong Hin Chuee v Citra Group Pte Ltd*, the High Court stated that subsequent conduct can be considered "if evidentially probative".[25] Similarly, in *Tienrui Design & Construction Pte Ltd v G & Y Trading and Manfacturing Pte Ltd*, the High Court allowed subsequent conduct to be admitted where it would help "to discern the objective intentions of the parties".[26] While these are undoubtedly restrictions on the admissibility of subsequent conduct, they are fairly easy to satisfy since it is really up to the court to assess the usefulness of such evidence in an instant case.

The third view is that subsequent conduct can only be admitted on the satisfaction of stringent conditions. These conditions differ from the limited restrictions discussed above in the sense that they depend on a defined set of objective criteria rather than a general criterion of "usefulness". Thus the High Court in *Sheng Siong Supermarket Pte Ltd v Carilla Pte Ltd*[27] rejected the plaintiff's attempt to admit the defendant's subsequent communication to interpret a lease agreement between the parties on the basis that the evidence constituted "both evidence of subsequent conduct and subjective intention". Such evidence, according to the court, would

> "tilt the balance towards uncertainty and thus does not constitute helpful evidence that goes towards proof of what the parties, from an objective viewpoint, ultimately agreed upon".[28]

In other words, the court understood *Zurich Insurance* to bar sebsequent subsequent conduct where such evidence showed the parties' subjective intention. In *HSBC Trustee (Singapore) Ltd v Lucky Realty Co Pte Ltd*, the High Court, after an extensive examination of the case law, surmised that the correct understanding post-*Zurich Insurance* is that there should be a heightened caution against the admission of subsequent conduct because they are likely to be "shaped towards self-serving ends with the benefit of hindsight, highly likely to be highly contentious and very often of only marginal assistance in construing a contract".[29] The court thus held that the Singapore approach is a "middle approach", which rejects a blanket exclusionary rule but imposes the *Zurich Insurance* requirements as

[22] *Goh Guan Chong v AspenTech Inc* [2009] 3 S.L.R.(R.) 590 at [57].
[23] *Goh Guan Chong v AspenTech* [2009] 3 S.L.R.(R.) 590 at [58].
[24] *Fico Sports Inc Pte Ltd v Thong Hup Gardens Pte Ltd* [2010] SGHC 237 at [60].
[25] *Leong Hin Chuee v Citra Group Pte Ltd* [2015] 2 S.L.R. 603 at [91].
[26] *Tienrui Design & Construction Pte Ltd v G & Y Trading and Manfacturing Pte Ltd* [2015] 5 S.L.R. 852 at [52].
[27] *Sheng Siong Supermarket Pte Ltd v Carilla Pte Ltd* [2011] 4 S.L.R. 1094.
[28] *Sheng Siong Supermarket v Carilla* [2011] 4 S.L.R. 1094 at [55].
[29] *HSBC Trustee (Singapore) Ltd v Lucky Realty Co Pte Ltd* [2015] 3 S.L.R. 885 at [50].

conditions to admission.[30] Similarly, in *BCBC Singapore Pte Ltd v PT Bayan Resources TBK*, the Singapore International Commercial Court,[31] a recently established division of the High Court, held that subsequent conduct is "unlikely to be admissible in many cases" owing to non-compliance with the *Zurich Insurance* requirements.[32] Indeed, on the facts, the court rejected the defendants' use of subsequent conduct because they were not relevant, reasonably available or related to a clear and obvious context.[33] One problem with this understanding of *Zurich Insurance* is of course that the Court of Appeal never actually imposed the *Zurich Insurance* requirements on the use of subsequent conduct. In fact, in *Ding Pei Zhen v Yap Son On*,[34] the High Court interestingly suggested that the *Zurich Insurance* requirements could never be satisfied in relation to subsequent conduct. According to the court, the requirement of reasonable availability to all contracting parties in *Zurich Insurance* refers to the extrinsic evidence being available *at the time of contracting*. Similarly, the court also pointed out that the requirement that there was a clear and obvious context refers to the context in which the contract was made. The court therefore thought that subsequent conduct could *never* satisfy the *Zurich Insurance* requirements and hence will *always* be inadmissible. However, the court did think that it is arguable that these requirements are not intended to be "limited temporally".[35]

The High Court's ambivalence about the admissibility of subsequent conduct has been matched by the Court of Appeal. There are some cases, such as *Sports Connection Pte Ltd v Deuter Sports GmbH*,[36] where the Court of Appeal admitted subsequent conduct. *Sports Connection* concerned the interpretation of a non-competition clause in a distributorship agreement entered between the parties. The Court of Appeal admitted an email sent after the conclusion of the distributorship agreement. Importantly, the court emphasised that it was not laying down a general principle given its statement in *Zurich Insurance* that the relevance of subsequent conduct requires detailed examination at a later time. Indeed, it stressed that subsequent conduct was unlikely to be admissible where

> "a party attempts to trawl through evidence in an attempt to favour its subjective interpretations of the contract and/ or where a party attempts to persuade the court to adopt a different interpretation from that suggested by the plain language of the contract".[37]

With this in mind, the court noted that the email

> "sheds valuable light on the *context* surrounding the intention of the parties when they entered into the Distributorship Agreement in general and agreed upon the Non-Competition Clause in particular".[38]

[30] *HSBC Trustee (Singapore) v Lucky Realty* [2015] 3 S.L.R. 885 at [50].
[31] See M. Yip, "The Resolution of Disputes before the Singapore International Commercial Court" (2016) 65 I.C.L.Q. 439.
[32] *BCBC Singapore v PT Bayan Resources* [2016] SGHC(I) 1 at [103].
[33] *BCBC Singapore v PT Bayan Resources* [2016] SGHC(I) 1 at [269].
[34] *Ding Pei Zhen v Yap Son On* [2015] 5 S.L.R. 911.
[35] *Ding Pei Zhen v Yap Son On* [2015] 5 S.L.R. 911 at [95].
[36] *Sports Connection Pte Ltd v Deuter Sports GmbH* [2009] 3 S.L.R.(R.) 883 (Sports Connection).
[37] *Sports Connection* [2009] 3 S.L.R.(R.) 883 at [70].
[38] *Sports Connection* [2009] 3 S.L.R.(R.) 883 at [71] (emphasis in original).

Although the court was at pains to stress that it was not laying down any general proposition, the practical effect of *Sports Connection* may well be that it had. By allowing subsequent conduct to be admitted where it illuminated the context rather than show the parties' subjective intentions or a meaning contrary to the contractual plain meaning, the court appeared to have sanctioned the admission of subsequent conduct in these circumstances.[39]

Notwithstanding this possible reading of *Sports Connection*, in the next case that it had to consider the admissibility of subsequent conduct, the Court of Appeal in *Straits Advisors Pte Ltd v Behringer Holdings (Pte) Ltd* implicitly maintained that it had not determinatively considered whether subsequent conduct could be admitted for contractual interpretation as a matter of law.[40] Later, the Court of Appeal in the important case of *Sembcorp Marine Ltd v PPL Holdings Pte Ltd*[41] appeared to restrict the more liberal approach towards contractual interpretation taken in *Zurich Insurance*. The court noted that the unrestrained admission of extrinsic material in the interpretation of contracts will cause uncertainty for parties and increase the time and cost of legal proceedings. Thus, it held that a "robust approach", characterised by the admission of prior negotiations and subsequent conduct, might lead to more fog and be ultimately counterproductive.[42] It was in this light that the court left for further consideration the issue of whether prior negotiations could be admitted under Singapore law for contractual interpretation.[43] The position after *Sembcorp Marine* thus is that, even though the admission of subsequent conduct (and prior negotiations) remains "unsettled", the pendumlum had swung against their admission. Indeed, the Court of Appeal in *Alwie Handoyo v Tjong Very Sumito* doubted whether the trial judge was correct in relying on subsequent conduct to interpret the contract concerned because such conduct was unconnected with the parties' intentions in the original contract.[44] This restrictive approach can also be seen in *Xia Zhengyan v Geng Changqing*,[45] which concerned the interpretation of a share transfer clause. The respondent argued that the appellant's subsequent failure to complain when certain shares were not transferred was admissible subsequent conduct showing that the parties did not intend those shares to be transferred. The Court of Appeal disagreed, holding that, even if there were no blanket prohibition against the admissibility of such evidence, the subsequent conduct would not be helpful because the appellant's failure to complain could simply be due to the ignorance of her legal right to the non-transferred shares.[46] Furthermore, in *YES F&B Group Pte Ltd v Soup Restaurant Singapore Pte Ltd*, the Court of Appeal disallowed the admission of subsequent conduct as being unhelpful on the basis that "the object of the interpretative exercise was to discern the parties' intentions at the time of entering into the contract", and that

[39] Indeed, in *Ashlock William Grover v SetClear Pte Ltd* [2012] 2 S.L.R. 625 at [23], the Court of Appeal rejected an attempt to admit subsequent conduct to interpret a contract on the ground that it "principally pertained to the Appellant's subjective belief as to the actual ambit of [the clause concerned]".

[40] *Straits Advisors Pte Ltd v Behringer Holdings (Pte) Ltd* [2010] 1 S.L.R. 760 at [17]: "In the circumstances, since the Termination Agreement is not even relevant conduct in the first place, we do not propose to discuss whether it is otherwise admissible subsequent conduct (in law) for the purposes of interpreting the Consultancy Agreement."

[41] *Sembcorp Marine Ltd v PPL Holdings Pte Ltd* [2013] 4 S.L.R. 193 (*Sembcorp Marine*).

[42] *Sembcorp Marine* [2013] 4 S.L.R. 193 at [72].

[43] *Sembcorp Marine* [2013] 4 S.L.R. 193 at [75].

[44] *Alwie Handoyo v Tjong Very Sumito* [2013] 4 S.L.R. 308 at [85].

[45] *Xia Zhengyan v Geng Changqing* [2015] 3 S.L.R. 732.

[46] *Xia Zhengyan v Geng Changqing* [2015] 3 S.L.R. 732 at [74].

reliance on such evidence is dangerous because "it can, with the benefit of hindsight, be shaped to suit each party's position".[47] If this were intended as a statement of principle, then it would apply to every instance that subsequent conduct is being admitted; it would, in other words, be difficult to conceive of a situation where subsequent conduct *can* be admitted to interpret a contract.

In sum, the Court of Appeal's approach towards subsequent conduct for contractual interpretation is thus more restrictive than the High Court's. In fact, it is difficult to find a recent case where the court has admitted subsequent conduct to interpret a contract. This restrictive approach is borne out by the Court of Appeal's latest statement on the admissibility of subsequent conduct in *Hewlett-Packard Singapore (Sales) Pte Ltd v Chin Shu Hwa Corinna*.[48] In this case, the court repeated its observations in *Soup Restaurant* that because contractual interpretation is concerned with the objective ascertainment of the parties' intentions at the time the contract was formed, subsequent conduct, "in so far as they reveal the subjective intention of the parties, will generally be irrelevant in this exercise".[49] Bringing us full circle to the English cases that laid down a blanket prohibition against subsequent conduct, the court then referred to *James Miller* and *Schuler* for the rationale why subsequent conduct has generally been precluded. However, the court then clarified that it was not endorsing a blanket prohibition on the use of subsequent conduct; any evidence can be admitted if it fulfils the *Zurich Insurance* requirements. Ultimately, the court repeated that whether subsequent conduct should be admitted to interpret contracts under Singapore law remained to be determined at a more appropriate time.

For present purposes, the following may be said about the use of subsequent conduct for contractual interpretation in Singapore. First, the High Court has advanced at least three different views post-*Zurich Insurance* on the use of subsequent conduct, revealing a lack of consensus on this important issue. Secondly, the Court of Appeal has maintained that there is no blanket prohibition against the use of subsequent conduct but that the issue remains an open one to be determined more conclusively in the future. Thirdly, notwithstanding its repeated emphasis that the issue is not concluded, the Court of Appeal's approach in recent cases hints at a highly restrictive approach in relation to subsequent conduct. Although formally constrained by the *Zurich Insurance* requirements, it seems exceedingly difficult to ever find that subsequent conduct has satisfied those requirements. This is because of the court's frequent allusion to the fact that subsequent conduct does not, logically, show the parties' intention at the time of formation and hence can hardly be "relevant" for the interpretative exercise. Moreover, the Court of Appeal has been very ready to find alternative explanations to account for the parties' subsequent conduct, rather than say that such conduct represented their understanding of the contract's objective meaning.

[47] *YES F&B Group Pte Ltd v Soup Restaurant Singapore Pte Ltd* [2015] 5 S.L.R. 1187 at [74] (*Soup Restaurant*).
[48] *Hewlett-Packard Singapore (Sales) Pte Ltd v Chin Shu Hwa Corinna* [2016] 2 S.L.R. 1083.
[49] *Hewlett-Packard Singapore v Chin Shu Hwa Corinna* [2016] 2 S.L.R. 1083 at [54]. The Court of Appeal further referred to its own decision in *Lian Hwee Choo Phebe v Maxz Universal Development Group Pte Ltd* [2009] 2 S.L.R.(R.) 624, where it endorsed the principle that a contract must generally be interpreted as at the date it was made and in light of the circumstances prevailing on the date (at [11]).

Contractual formation

In contrast to the more restrictive, even impossible, use of subsequent conduct to interpret contracts, the Singapore courts have been more willing to use subsequent conduct to decide formation issues, such as the existence or nature of a contract, or the nature of its terms. Whether this is consistent with the more restrictive use of subsequent conduct in contractual interpretation will be discussed after considering these uses in brief.

Ascertaining the existence of a contract

First of all, the Singapore courts have used subsequent conduct to ascertain the existence of a contract. The High Court has recently stated in *G-Fuel Pte Ltd v Gulf Petrochem Pte Ltd* that[50]:

> "Where the issue before the court is whether or not a contract was concluded between the parties, subsequent conduct may be relevant. In considering such conduct, the court is not looking at past conduct to construe the terms of the contract, a matter dealt with extensively by the Court of Appeal in [*Sembcorp Marine*], or to see whether new terms may be introduced into the contract."

Thus, in *Hongkong & Shanghai Banking Corp Ltd v Jurong Engineering Ltd*,[51] one of the issues was whether the first defendant had entered into a compromise agreement with the plaintiff that set out a repayment schedule in settlement of a debt. The High Court held that the parties' subsequent conduct was admissible to determine if the parties' exchange of letters at the material time constituted valid offer and acceptance.[52] The court took into account the fact that, after the parties exchanged letters, the plaintiffs, after initially sending two letters, failed for the next two years to remind the first defendant of its liability under any compromise agreement. There was also no record of the plaintiffs contemplating any legal action against the first defendant in this period for non-payment. Indeed, even when the plaintiffs sent letters of demand in respect of the original debt, they made no mention of the compromise. Accordingly, the court found that the plaintiffs had never treated the first defendant's original letter as an offer to undertake any repayment liability, and therefore never accepted it. Similarly, in *Midlink Development Pte Ltd v The Stansfield Group Pte Ltd*,[53] the High Court had to decide whether there existed an oral lease agreement that arose between the plaintiff and defendant. The court considered that the defendant's subsequent conduct of paying the reduced rent pursuant to the alleged agreement was relevant in determing its existence. Indeed, the court found that the defendant was fully aware that the plaintiff believed a binding agreement had been reached, and was content to go along with that belief. The court therefore found that the parties had, in fact, concluded an oral lease agreement. Finally, and in a related fashion, the High Court in *Bank of China Ltd (Singapore Branch) v Huang Ziqiang* used subsequent conduct to ascertain the time at which a contract, in that case a guarantee, was created.[54]

[50] *G-Fuel Pte Ltd v Gulf Petrochem Pte Ltd* [2016] SGHC 62 at [50].
[51] *Hongkong & Shanghai Banking Corp Ltd v Jurong Engineering Ltd* [2000] 1 S.L.R.(R.) 204.
[52] *Hongkong & Shanghai Banking Corp v Jurong Engineering* [2000] 1 S.L.R.(R.) 204 at [78].
[53] *Midlink Development Pte Ltd v The Stansfield Group Pte Ltd* [2004] 4 S.L.R.(R.) 258.
[54] *Bank of China Ltd (Singapore Branch) v Huang Ziqiang* [2004] SGHC 245 at [63].

The court relied on the defendant's subsequent instruction to his lawyers in another action that the guarantee took effect on a particular date to find that it had, in fact, taken effect on that date. The defendant's conduct, in the court's words, "represented and confirmed his mindset that the Guarantee was executed on [the particular date]".[55]

Apart from these cases, which used subsequent conduct to find that a contract had been formed, the High Court in *Chor Pee & Partners v Wee Soon Kim Anthony*[56] and *United Fiber System Ltd v China National Machinery & Equipment Import & Export Corp*[57] used subsequent conduct to find that a contract had *not* been formed.[58] In the former case, the court found that the parties had not entered into any agreement on legal fees since the respondent emphatically disagreed with the petitioner's subsequent letter seeking to confirm that certain sums were payable.[59] In *United Fiber*, the parties' continued negotiations and the plaintiff's failure to assert that there had been a breach of the alleged agreement led to the court to find that there was no agreement. Similarly, the Court of Appeal in *Ng Chee Chuan v Ng Ai Tee*[60] used subsequent conduct to decide that there was no oral agreement between the parties for the recovery of certain moneys. The court found that the respondent's failure to sue on its legal rights pursuant to the alleged oral agreement supported the finding that no such agreement existed.[61]

Notwithstanding the cases that have used subsequent conduct to determine the existence (or non-existence) of a contract, both the High Court *and* the Court of Appeal have doubted whether this is actually permitted as a matter of law. For example, the High Court in *ARS v ART* observed that

> "given the complex relationships between human beings … conduct can be explained by a number of reasons which does not have only one explanation or there may be varying degrees of weight pointing to one conclusion".[62]

The court's point is that the parties' subsequent conduct cannot, in most cases, point conclusively to the existence of a contract, since such conduct could be explained by other reasons.[63] More recently, the High Court in *Grains and Industrial Products Trading Pte Ltd v Bank of India*[64] also disagreed with the plaintiff's attempt to use subsequent conduct to infer the existence of a contract.

What is conceptually significant about these cases is that the High Court drew the conceptual connection between the use of subsequent conduct to interpret a contract and to ascertain the existence of a contract. Thus, the court in *ARS* questioned whether the Court of Appeal's statement in *Zurich Insurance* that the

[55] *Bank of China v Huang Ziqiang* [2004] SGHC 245 at [66]. See also the Court of Appeal decision of *Chin Siew Seng v Quah Hun Kok Francis* [2010] SGCA 44 at [28].

[56] *Chor Pee & Partners v Wee Soon Kim Anthony* [2005] 3 S.L.R.(R.) 433.

[57] *United Fiber System Ltd v China National Machinery & Equipment Import & Export Corp* [2011] 2 S.L.R. 1021 (*United Fiber*).

[58] See also *Aragorn Assets Ltd v Bright Ideas Ltd* [1999] SGHC 235 at [70]–[71].

[59] *Chor Pee v Wee Soon Kim Anthony* [2005] 3 S.L.R.(R.) 433 at [53].

[60] *Ng Chee Chuan v Ng Ai Tee* [2009] 2 S.L.R.(R.) 918.

[61] See also *Econ Corp Ltd v So Say Cheong Pte Ltd* [2004] SGHC 234 at [31].

[62] *ARS v ART* [2015] SGHC 78 at [90].

[63] To be fair, however, in many of the cases where subsequent conduct was used to support a finding or negation of a contract, subsequent conduct was used *together* with other evidence to support the court's eventual finding—although this does mean the implicit acceptance that subsequent conduct could be admitted as a matter of law in the first place.

[64] *Grains and Industrial Products Trading Pte Ltd v Bank of India* [2015] 1 S.L.R. 1213 (*Grains*).

admissibility of subsequent conduct for contractual interpretation remained a controversial point should apply consistently to the use of such evidence to find the existence of a contract. In *Grains*, the court acknowledged that *Zurich Insurance* concerned contractual interpretation, but thought that

> "the principles enunciated there on the relevance of evidence of subsequent conduct can be properly applied to the issue of contract *formation*, which, like contractual interpretation, turns on the intentions of the contracting parties objectively ascertained".[65]

Based on this, the court held that subsequent conduct cannot be used "to add to, vary or contradict the terms of the oral contract once the contract is formed".[66] In fact, even earlier, in *Sundercan Ltd v Salzman Anthony David*[67] and *Bridgeman Pte Ltd v Dukim International Pte Ltd*,[68] the High Court had said that for consistency with *Zurich Insurance*, evidence of subsequent conduct may be irrelevant to the determination of contractual terms at the time of formation.[69] Indeed, in *Bridgeman*, the court reasoned that subsequent conduct would hence be "inadmissible as direct evidence of contractual terms".[70]

Even more important may be the Court of Appeal's view in *Fairview Developments Pte Ltd v Ong & Ong Pte Ltd* that subsequent conduct "would not affect the coming into existence of the contract … at an *earlier* point in time".[71] To be fair, the court is not actually saying that subsequent conduct cannot be used to *support* the finding that there had been a contract. However, practically speaking, the use of subsequent conduct as part of the evidence to infer the existence of a contract is close to using it to "affect" the coming into existence of the contract. It may thus be possible to read the Court of Appeal in *Fairview Developments* to be saying that subsequent conduct can no longer be used to determine the existence of a contract under Singapore law, although the better view is that the issue remains, as with the use of subsequent conduct for interpretation, an open one under Singapore law. This is particularly so in light of the High Court's uneasiness in some cases regarding the apparent inconsistency between the use of such evidence for interpretation and formation issues.

Before ending this section, it may be noted for completeness that "subsequent" conduct can certainly be used to infer the existence of a *new* contract.[72] This is different from the preceding discussion as the "subsequent" conduct here is used to infer the existence of an entirely *new* contract *after* the relevant conduct, rather

[65] *Grains* [2015] 1 S.L.R. 1213 at [24].

[66] *Grains* [2015] 1 S.L.R. 1213 at [27]. However, even though the High Court in *Grains* purported to agree that subsequent conduct should not be used as direct evidence of contractual terms, it then, in apparent contradiction to what it said, went on to say that at some level at least, the existence of an oral agreement can be inferred from the conduct of the parties, both before and after the time the oral contract was allegedly concluded (at [25]).

[67] *Sundercan Ltd v Salzman Anthony David* [2010] SGHC 92 (*Sundercan*).

[68] *Bridgeman Pte Ltd v Dukim International Pte Ltd* [2013] SGHC 220 (*Bridgeman*).

[69] *Sundercan* [2010] SGHC 92 at [27]; and *Bridgeman* [2013] SGHC 220 at [14]. However, in both cases, the court was prepared to assume that subsequent conduct could be used as direct proof of contractual terms in the absence of complete arguments by the parties.

[70] *Bridgeman* [2013] SGHC 220 at [14]. However, the court was prepared to assume that subsequent conduct could be used as direct proof of contractual terms in the absence of complete arguments by the parties.

[71] Although subsequent conduct may be relevant to the finding of a later variation: see *Fairview Developments Pte Ltd v Ong & Ong Pte Ltd* [2014] 2 S.L.R. 318 (*Fairview Developments*) at [78] (emphasis in original).

[72] See *Great Eastern Life Assurance Co Ltd v Ng Hui Lip* [1983–1984] S.L.R.(R.) 400; *Ho Dwan Tiauw v Banque Nationale de Paris* [1993] 1 S.L.R.(R.) 524 at [23]; and *Star-Trans Fat East Pte Ltd v Norske-Tech Ltd* [1996] 2 S.L.R.(R.) 196.

than being used to infer the existence of a contract preceding it. However, the High Court had interestingly noted in *Sim Tony v Lim Ah Ghee* that

> "the fact that a person considers himself to have made a binding agreement and that his subsequent conduct was consistent with there being an agreement is not in itself evidence of anything".[73]

Determining the nature of a contract

Apart from ascertaining the existence of a contract, at least one Singapore case has used subsequent conduct to determine the *nature* of a contract. In *EC Investment Holding Pte Ltd v Ridout Residence Pte Ltd*, the High Court, while acknowledging that the admissibility of subsequent conduct for contractual interpretation remained an unsettled issue following *Zurich Insurance*, nonetheless held that subsequent conduct can be admitted to see how the parties "implemented or treated the contract or contracts".[74] One of the issues in *EC Investment* was whether a purported option to purchase was really a loan. The court noted that the subsequent conduct of the plaintiff and its lawyer was "quite at odds" with the contract being a true sale and purchase of property. In particular, the plaintiff and its lawyer were silent for some two months even though the defendant's lawyer did not respond to their exercise of the option. This, to the court, contradicted the plaintiff's assertion that it was "truly only interested to purchase the Property".[75] Moreover, the plaintiff did not offer any evidence of any serious steps to obtain financing for the purchase. The court thus found that the true nature of the agreement between the parties was a loan as opposed to an agreement to purchase the property concerned. While this finding was overturned on appeal,[76] the Court of Appeal did not disagree with the High Court's statement of principle that it could rely on subsequent conduct to evaluate the true nature of the agreement between the parties. Indeed, the Court of Appeal recognised that "the court is not prohibited from evaluating evidence *other than* the transaction documents … to determine the true nature of a transaction".[77]

Determining specific terms of a contract

Apart from broadly determining the existence or nature of a contract, the Singapore courts have used subsequent conduct more specifically to "determine the specific terms" of a contract. In *Yap Cheng Koon t/a Yap Cheng Koon Labour Services v His Intermedia Marketing (S) Pte Ltd*, the High Court stated that there was no reason why the parties' subsequent conduct could not be used to decide whether "the contract in fact contained a particular term".[78] An example of this use of subsequent conduct is the determination of the proper parties to a contract. In *Metro Pte Ltd v Wormald Security (SEA) Pte Ltd*,[79] the defendant contracted to

[73] *Sim Tony v Lim Ah Ghee* [1994] 2 S.L.R.(R.) 910 at [58].
[74] *EC Investment Holding Pte Ltd v Ridout Residence Pte Ltd* [2011] 1 S.L.R. 232 at [79] (*EC Investment*).
[75] *EC Investment* [2011] 1 S.L.R. 232 at [82].
[76] *EC Investment Holding Pte Ltd v Ridout Residence Pte Ltd* [2012] 1 S.L.R. 32.
[77] *EC Investment* [2012] 1 S.L.R. 32 at [30].
[78] *Yap Cheng Koon t/a Yap Cheng Koon Labour Services v His Intermedia Marketing (S) Pte Ltd* [1995] SGHC 74.
[79] *Metro Pte Ltd v Wormald Security (SEA) Pte Ltd* [1981–1982] S.L.R.(R.) 126 (*Metro*).

install and maintain an alarm system at the plaintiffs' store. The contract provided that the defendant shall alert the "customer or such person as the customer has nominated" should the alarm go off. When the alarm did go off and the defendant failed to alert the plaintiffs, the plaintiffs sued. One of the issues was whether the second plaintiff, who owned the store but obtained its goods from the first plaintiff, was a "customer" referred to in the clause regarding the alert. The issue arose because an executive from both plaintiffs' parent company had only specified the first plaintiff as the "customer" in the contract concerned. The High Court relied on the parties' subsequent conduct to find that the second plaintiff was indeed a "customer" under the relevant clause. After the contract was signed, the second plaintiff had written a letter to the defendant nominating the persons who should be alerted in the event the alarm went off. Since the defendant did not object, the court found that it had "tacitly accepted both plaintiffs as 'customers' of theirs under the contract".[80]

Such use of subsequent conduct can also be seen in the High Court case of *Abundance Development Pte Ltd v Absolut Events & Marketing Pte Ltd*.[81] The issue in the case was whether a licensing agreement, which the parties submitted was partly oral and partly written, for the use of an atrium extended to an area known as "lot 5". The court held that subsequent conduct may be used to ascertain "the terms of an oral and only partially expressed agreement", referring to some English cases in support and in apparent distinction from using such conduct to interpret a contract.[82] Yet, the court then acknowledged that the case turned on the "contrasting interpretations" offered by the parties.[83] The court found that the defendant's subsequent conduct in not requesting for the use of lot 5 supported its finding that the licensing agreement did not extent to lot 5. The court's inconsistent characterisation of the issue as the "ascertainment of terms" and the "interpretation" surely shows that the two are closely related, if not the same.

Variation or estoppel

Finally, apart from using subsequent conduct for interpretation and formation issues, the Singapore courts recognise that subsequent conduct can be used to infer the variation of the original contract or to support an estoppel. Thus, the High Court held in *Estate of Seow Khoon Seng, deceased v Pacific Century Regional Developments Ltd* that while subsequent conduct could not be used to construe a written agreement (at the time of its decision, at least), it could be looked at "when variation or estoppel is an issue".[84] Some 10 years later, the Court of Appeal reaffirmed these trite principles in *Singapore Telecommunications Ltd v Starhub Cable Vision Ltd*, where it said that although subsequent conduct could not be taken into account as an aid to interpretation, it may be "sufficient to show that

[80] *Metro* [1981–1982] S.L.R.(R.) 126 at [18].

[81] *Abundance Development Pte Ltd v Absolut Events & Marketing Pte Ltd* [2009] SGHC 198.

[82] *Ferguson v Dawson & Partners (Contractors) Ltd* [1976] 1 W.L.R. 1213 CA (Civ Div); *Wilson v Maynard Shipbuilding Consultants A B* [1978] Q.B. 665 CA (Civ Div); *Mears v Safecar* [1983] Q.B. 54 at 77; and *Carmichael v National Power Plc* [1999] 1 W.L.R. 2042 at 2050–2051.

[83] *Abundance Development* [2009] SGHC 198 at [1].

[84] *Estate of Seow Khoon Seng* [1997] 1 S.L.R.(R.) 1 at [27]. See also *Kredietbank v Sinotani Pacific* [1999] 1 S.L.R.(R.) 274 at [113]; *Panwah Steel v Koh Brothers* [2006] 1 S.L.R.(R.) 788 at [8]; and *Sherridon Exim Pte Ltd v India International Insurance Pte Ltd* [2001] SGHC 93 at [41].

[the parties] made a new contract or to constitute the basis of an estoppel".[85] Indeed, as for estoppel by convention, the High Court in *Hilborne v Singapore Island Country Club*[86] accepted Denning LJ's pronouncement in *Amalgamated Investment & Property Co Ltd v Texas Commerce International Bank Ltd* to the effect that if parties had by their course of dealing adopted a conventional basis for the governance of their relationship, then they will be bound by it.[87]

What are the inconsistencies in the uses of subsequent conduct in Singapore contract law?

What is "inconsistency" in the law?

Having now described the uses of subsequent conduct in Singapore contract law, let us now consider whether there are any inconsistencies in the uses. The need for the consistent application of legal doctrine is well accepted. Professor Robertson has written that courts are "constrained by the formal legal goal of ensuring consistency between related principles and related bodies of law".[88] This is said to ensure "doctrinal stability" and also ensures certainty in the rules that govern commerce.[89] Thus, Kirby J has said in *Cattanach v Melchior* that judges should apply the law "in ways that are logically reasoned and shown to be a consistent development of past decisional law" and "have no authority to adopt arbitrary departures from basic doctrine".[90] Consistency in the law is therefore pursued to ensure logic and certainty in the law; this does not appear to be a controversial point.

In ascertaining whether there is consistency, there is a need to consider what the "basic doctrine" is and to ask whether the further development or application of that doctrine is a logical extension of it. What then is the "basic doctrine" in the use of subsequent conduct for the uses of subsequent conduct that we have discussed? In order to answer this question, we need first to be clear about what we mean by "subsequent" conduct. This is because the sense in which "subsequent" is used may belie a different basic doctrine.

The basic distinction—"subsequent" relative to which point in time?

"Subsequent" is a measure of a point in time relative to some other point. There are two possible points of reference. The first is the conduct relative to the original contract (whether real or alleged). There is no real controversy here in the use of the expression "subsequent conduct": "subsequent" here means conduct *after* the formation of the original contract (whether real or alleged). Thus, conduct is always "subsequent" to the original contract whether it is being used to interpret that

[85] *Singapore Telecommunications Ltd v Starhub Cable Vision Ltd* [2006] 2 S.L.R.(R.) 195 at [26] (*Singapore Telecommunications*).

[86] *Hilborne v Singapore Island Country Club* [1996] 1 S.L.R.(R.) 654. See also *Mohamed Bassatne v Rifaat El Gohary* [2004] SGHC 63 at [118].

[87] *Amalgamated Investment & Property Co Ltd v Texas Commerce International Bank Ltd* [1981] 3 All E.R. 577 CA (Civ Div) at 583–584.

[88] A. Robertson, "Constraints on Policy-Based Reasoning" in A. Robertson and H.W. Tang (eds), *The Goals of Private Law* (Oxford: Hart Publishing, 2009), p.271.

[89] Robertson, "Constraints on Policy-Based Reasoning" in *The Goals of Private Law* (2009), p.271.

[90] *Cattanach v Melchior* (2003) 215 C.L.R. 1.

contract, determine the existence or nature of that contract, or to infer the variation of that contract, the formation of a new contract or the arising of an estoppel. The constant point of reference in all these uses remains the original contract.

The second point of reference is the time at which the fact, which subsequent conduct is being used to prove, arises. For example, if subsequent conduct is being used to interpret the meaning of the original contract, the conduct will certainly be "subsequent" in time to the fact that is being proved, in this case, the parties' intention at the point of contract formation. Likewise, the conduct will be "subsequent" if it is being used to prove the existence or nature of the original contract, or its specific terms. Here, as with the former example, the conduct takes place after the contract's form, nature or terms had come into being. In contrast, conduct is not "subsequent" by this point of reference if the conduct is being used to determine a fact that occurred *after* the conduct had taken place. For example, if "subsequent" conduct is used to determine whether the original contract had been varied, it is generally being used to decide a fact occurring *after* the conduct had taken place. This is because the variation would have arisen in most cases *as a result of* the "subsequent" conduct, rather than before it. Another example is the use of "subsequent" conduct to decide whether a new contract had been formed. While it is of course possible that the new contract was formed in between the formation of the original contract and the subsequent conduct, in most cases it would have been formed after the so-called "subsequent" conduct. Thus in these cases the "subsequent" conduct is being used to infer the existence of a *new* contract that had formed *as a result* or at least *after* the conduct.

Out of the two points of reference, it is the second that is the more important. This is because important considerations of principle and policy arise depending on *what* the subsequent conduct is being used to prove, rather than whether the subsequent conduct had arisen after the original contract. If so, we may at the outset distinguish the instances where subsequent conduct is used to establish a fact *prior* in time, and those where subsequent conduct is used to establish a fact *later* in time. Difficult questions of principle and policy arise mainly in respect of the former situation rather than the latter. The result is that, generally speaking, it is much harder to use subsequent conduct to prove a prior fact than it is to prove a later one. Take for example the use of subsequent conduct to prove variation or estoppel: the courts routinely admit subsequent conduct for this purpose and distinguish, as the Court of Appeal did in *Singapore Telecommunications*, between subsequent conduct which cannot be taken into account as an aid to interpretation, and such conduct that is "sufficient to show that [the parties] made a new contract or to constitute the basis of an estoppel".[91] Similarly, although the English courts have steadfastly refused to admit subsequent conduct to interpret contracts, they have routinely done so to find variation or estoppel. The implicit distinction that is drawn is between using subsequent conduct to prove a prior fact and using it to prove a latter one.

The inconsistency in the ease with which subsequent conduct is used can be explained by two principal reasons. First, when subsequent conduct is being used to determine a fact later in time, there is an easier causal connection both in terms of time and logic. All things being equal, it will be more natural to find that a fact

[91] *Singapore Telecommunications* [2006] 2 S.L.R.(R.) 195 at [26] .

is the result of something, rather than use a later conduct to infer the occurrence of a fact prior in time. Secondly, there are few policy reasons that preclude reference to conduct occurring prior in time to establish a later fact. Apart from the quality of the evidence, there is less concern of subjective manipulation of the evidence to achieve a certain desired outcome. There is also less uncertainty, since the meaning of the contract will not be unduly affected by events happening later in time. For these reasons, this article will put to one side the situations where "subsequent" conduct is being to establish a fact later in time. This will mainly include such conduct being used to show variation of the original contract, an estoppel or the formation of a new contract. These situations form a broad category of their own and do not give rise to the same issues of principle or policy as the situations where subsequent conduct is being used to prove a fact occurring prior in time. If there is any inconsistency in the use of subsequent conduct in these two broad categories, it can be explained on the basis that subsequent conduct is being used for two very different purposes.

Inconsistency in the use of subsequent conduct to prove a fact prior in time

We are therefore left with the situations where subsequent conduct is truly "subsequent" *both* in terms of the occurrence of the conduct relative to the original contract, *and* the fact that the conduct is meant to prove. The main situations here concern the use of subsequent conduct to interpret a contract, to ascertain the existence or nature of a contract, and to determine the existence of specific terms in a contract. It will be evident from the discussion above that subsequent conduct is not used consistently across these uses. For example, while it is difficult to admit subsequent conduct to interpret a contract, subsequent conduct is more easily used to infer the existence of a contract, ascertain its nature or determine the existence of specific terms. While there have been cases which have questioned the correctness of this inconsistency, the majority of Singapore cases have tended to use subsequent conduct inconsistently where such conduct is being used to prove a fact prior in time as a matter of contract law.

Argument for consistent use of subsequent conduct to prove a fact prior in time based on the same underlying basic use

The inconsistent use of subsequent conduct to prove a fact prior in time is not justified because it is not logical and does not promote certainty to have different applications of a basic doctrine. This can be illustrated with an anchor situation with which to compare the rest. Let us take the use of subsequent conduct to interpret a contract. The determination of the nature of a contract cuts close to the interpretation of its terms. In that regard, it is trite law that contractual interpretation is concerned with an objective ascertainment of the parties' intention. Indeed, objectivity has been described by the Court of Appeal in *Zurich Insurance* as "the cornerstone of the theory of contract and permeates our entire approach to contractual interpretation".[92] So at bottom, when subsequent conduct is used to

[92] *Zurich Insurance* [2008] 3 S.L.R.(R.) 447 at [69].

interpret a contract, it is being used to ascertain the parties' intentions prior in time. Put more specifically, the fact that the subsequent conduct is being used to prove is the fact of the parties' specific intention in relation to a particular term in the contract.

If expressed this way, is the use of subsequent conduct to interpret a contract any different from the other situations where such conduct is being used to prove a fact prior in time under Singapore contract law? Let us take the Singapore courts' use of subsequent conduct to "determine the specific terms" of a contract. At first blush, there appears to be no distinction between the ascertaining the "specific terms of a term" and the "meaning of a term in the contract". Let us take the example of *Metro* that was discussed earlier. It will be recalled that the High Court relied on the parties' subsequent conduct to find that the second plaintiff was indeed a "customer" under the relevant clause. Although the High Court phrased its approach in *Metro* as determining the specific terms of a contract, this is substantively an interpretation of the contract. Indeed, if a court is willing to use subsequent conduct to ascertain the proper parties to a contract by "ascertaining the specific terms of a contract", why should there be a prohibition against subsequent conduct if the question of who the proper parties are depends on the interpretation of a term within that contract? In both instances, the court is concerned with ascertaining the parties' intention; indeed, it is suggested that the intention sought to be proved is the same in both situations, that is, the parties' specific intention in relation to a particular term in the contract.

Moving on then to the use of subsequent conduct to prove the nature of a contract, it is submitted that this is merely a broader instance of interpreting the contract. Indeed, a contract is made up of its constituent terms and the ascertainment of its nature is surely quite similar to the interpretation of its specific terms. Take for example the High Court's use of subsequent conduct in *EC Investment* to show that the agreement between the parties was in fact a loan and not an option to purchase. Suppose that the question of the nature of the transaction boiled down to the interpretation of a clause in the contract, say for example a clause that contemplated the transfer of the property for a price that could be returned with interest for the property later on, surely there would be no difference between using subsequent conduct more specifically here to discern if the parties intended this clause to be a sale of the property or a loan secured by the property?

Finally, the distinction between using subsequent conduct to interpret a contract and to ascertain the existence of one has caused much disquiet among the Singapore courts. It is not hard to see why. At its broadest, the existence of a contract is as dependent on ascertaining the parties' intentions as it is on interpreting the contract. This explains why the High Court was uncomfortable with a more liberal admission of subsequent conduct to ascertain the existence of a contract in *ARS*, *Grains*, *Sundercan* and *Bridgeman*, as compared with the more restrictive approach towards the contractual interpretation. There was tacit acknowledgement in all those cases that the two are ultimately concerned with the ascertainment of the parties' intentions, and there should therefore be consistency in how subsequent conduct is used between them.

Accordingly, if we accept that the substantive purpose of using subsequent conduct to prove a prior fact under Singapore contract law is to prove the parties' intentions, even if the content of those intentions differ, then it is difficult to see why there should be inconsistencies between these different uses.[93] The next part of this article will further explain why, apart from the reasons already canvassed, there needs to be a consistent use of subsequent conduct to prove a fact prior in time.

Arguments for a (consistent) expansive use of subsequent conduct to prove a fact prior in time based on principle and policy

So far we have seen that because subsequent conduct is being used to prove the parties' intention, whether it is being used to interpret a contract, ascertain the existence/nature of a contract or determine its specific terms, it ought to be *consistently* used across these uses. It would be logically inconsistent for subsequent conduct to be freely used to infer the existence of a contract but restrictively used to interpret a contract. In all cases, the ultimate use of subsequent conduct is the same and hence the degree of its use ought also to be the same. However, this does not tell us *what* that consistent approach ought to be. It might well be that a restrictive approach is to be applied *consistently* across all uses.

This section argues that that subsequent conduct should be used expansively whenever it is being used to prove a fact prior in time. Two points will be advanced. First, as a matter of principle, Singapore evidence law mandates an expansive use of subsequent conduct to prove a fact prior in time. Secondly, even if this first point is wrong, it will be argued that an expansive use of subsequent conduct is aptly justified by policy reasons.

Evidence Act mandates expansive admission of subsequent conduct

A question of admissibility?

The relevance of the Singapore Evidence Act[94] depends on whether the use of subsequent conduct in contract is really one of admissibility under the law of evidence. If the rules governing the use of subsequent conduct in Singapore contract law are substantive rules of contract, then the evidentiary rules found in the EA must yield to them since procedural rules are always secondary to substantive ones. However, it is more likely that the rules governing the use of subsequent conduct are part of the law of evidence and not of contract. Indeed, the Court of Appeal in *Sembcorp Marine* said that the admissibility of extrinsic evidence in Singapore is "primarily statutory in the form of the EA".[95] The court further held that the EA only governs the admissibility of evidence but does not prescribe the rules of contractual construction.[96] Accordingly, the court drew a distinction between the

[93] See also D.W. McLauchlan, "Contract Formation, Contract Interpretation, and Subsequent Conduct" (2006) 25 U.Q.L.J. 77, who reaches a similar conclusion.
[94] Singapore Evidence Act Cap.97, 1997 Rev. Ed. (EA).
[95] *Sembcorp Marine* [2013] 4 S.L.R. 193 at [39].
[96] *Sembcorp Marine* [2013] 4 S.L.R. 193 at [40].

> "rules of contract law, viz, the substantive law which determines rights and liabilities, and rules of evidence, viz, the procedural law which determines what and how facts may be proved".[97]

The court nonetheless recognised that one set of rules may affect the other; the rules of evidence may, for instance, affect the *application* of specific rules of contractual interpretation even if they do not prescribe *how* a contract is to be interpreted.[98]

The distinction between a rule of contract and a rule of evidence may be illustrated by the principles of contractual interpretation. As the Court of Appeal said in *Sembcorp Marine*, the contextual approach in contractual interpretation is a rule of contract. That approach to interpreting a contract leads to a determination of the parties' rights and liabilities. However, what evidence is admissible so as to form the "context" is a question of evidence. The "context" in the contextual approach may be extended or limited by the rules of evidence. That is not to say that the rules of evidence override the contextual approach, but they do affect the *application* of the contextual approach, a rule of contract. For present purposes, if we take the view that subsequent conduct is being used to determine the parties' intention prior in time, the relevant rule of contract is that the parties' intention are crystallised at the point of formation. The crystallisation of their intention thus determines their respective rights and liabilities arising from the contract. But what is admissible to *prove* this intention is a rule of evidence. Thus, while the admissibility of subsequent conduct to prove this intention does not *prescribe* when the parties' intention are taken in law to have crystallised, the admissibility of subsequent conduct can undoubtedly affect the *application* of this rule of contract by controlling the facts admissible to prove such intention.

If the use of subsequent conduct is really a question of admissibility and a rule of evidence, it is necessary to start our discussion by referring to the EA, which governs the range of evidence admissible, of which subsequent conduct is an example. Unlike English evidence law, which prescribes a broad criterion of "relevance" to determine the admissibility of evidence, the EA operates more specifically through several related provisions. The general provision governing admission of evidence is s.5, which provides as follows:

> "Evidence may be given in any suit or proceeding of the existence or non-existence of every fact in issue and of such other facts as are hereinafter declared to be relevant, and of no others."

As Professor Pinsler explains, there are two fundamental categories of admissible evidence under the EA, namely "facts in issue", which are primary facts that concern liability and non-liability, and "relevant facts", which are the related facts from which the facts in issue can be inferred.[99] These are evident from s.2 of the EA, which defines a "fact in issue" as

[97] *Sembcorp Marine* [2013] 4 S.L.R. 193 at [43].

[98] *Sembcorp Marine* [2013] 4 S.L.R. 193 at [40]. See also HSBC Trustee (Singapore) Ltd v Lucky Realty Co Pte Ltd [2015] 3 S.L.R. 885 at [52].

[99] J. Pinsler SC, *Evidence and the Litigation Process*, 3rd edn (LexisNexis, 2010), p.33.

> "any fact from which either by itself or in connection with other facts the existence, non-existence, nature or extent of any right, liability or disability asserted or denied in any suit or proceeding necessarily follows".

It also defines a fact as "relevant" to another "when the one is connected with the other in any of the ways referred to in the provisions of this Act relating to the relevancy of facts". Thus, by way of illustration, if A brings an action for breach of contract against B, and B denies that the contract was ever formed, then the facts in issue will include whether the contract was ever formed, whether it was breached, and the loss suffered by A.[100] A relevant fact may be B's conduct after the contract was formed, from which the fact in issue concerning the formation of the contract may be inferred.

Of course, direct evidence of a fact in issue is always superior to circumstantial evidence of the same fact. However, direct evidence is not always available and hence facts in issue tend to be proved by relying on "relevant facts".[101]Sections 6 to 16 of the EA define exhaustively what counts as a "relevant fact" for the purpose of admission under s.5. Thus, as the Court of Appeal explained in *Lee Chez Kee v PP*, the EA adopts "an inclusionary approach, stating what may be admitted in evidence".[102] The key question for admissibility under Singapore evidence law is whether the evidence sought to be admitted satisfies any of the definitions of legal (not logical) relevancy as defined by the EA; if so, that fact is made admissible by s.5.[103]

Understood this way, there are a few questions that need to be answered as a matter of evidence law in our discussion relating to subsequent conduct. First of all, although subsequent conduct will almost never be direct evidence of a fact in issue, is it nonetheless a "relevant" fact under the terms of the EA, rendering it admissible? Secondly, if subsequent conduct is indeed admissible, can the courts formulate common law rules restricting its admissibility?

Evidence Act allows the admission of subsequent conduct

The EA does in fact allow the admission of subsequent conduct. Section 8(2) provides as follows:

> "The conduct of any party or of any agent to any party to any suit or proceeding in reference to such suit or proceeding or in reference to any fact in issue therein or relevant thereto, and the conduct of any person an offence against whom is the subject of any proceeding, is relevant if such conduct influences or is influenced by any fact in issue or relevant fact, and whether it was previous or subsequent thereto."

Explanation 1 to s.8 provides that conduct does not include statements unless they accompany and explain acts. As Professor Pinsler explains, s.8(2)

[100] Pinsler, *Evidence and the Litigation Process* (2010), p.34.
[101] Pinsler, *Evidence and the Litigation Process* (2010), p.35.
[102] Evidence Act Cap. 97, 1997 Rev. Ed.
[103] Evidence Act Cap. 97, 1997 Rev. Ed.

> "admits evidence of the conduct of a party or his agent — including the conduct of a person accused in the proceedings — which is influenced by, or influences, any fact in issue or relevant issue".[104]

The courts are conscious that this is merely an inference and so do not regard evidence admitted under s.8(2) as conclusive. As the High Court stated in *Lewis Christine v PP*, a criminal case, "conduct subsequent to the offence, without more, is not conclusive of prior guilt".[105]

Indeed, most of the cases that have admitted evidence under s.8(2) have concerned criminal matters. Thus, in *Krishna Jayaram v PP*,[106] it was held that an accused person's conduct of arrangement for a meeting with a person to change his statement to the prosecuting authority was admissible as relevant subsequent conduct. In *Chanderasekaran v PP*,[107] a person accused of obtaining money in a dishonest fashion giving false explanations that he obtained the money honestly may be admissible subsequent conduct influenced by the facts in issue. Similarly, a person's attempt to escape from custody has been admitted as a relevant fact to support the inference that she had commtitted theft.[108] All of these cases have involved the admission of subsequent conduct to prove a prior fact, here, the accused's guilt in relation to the alleged offence.

Although used primarily in criminal cases, it is clear that s.8(2) can apply to civil cases as well. Indeed, the section does not limit its application to only criminal cases but refers to the conduct of any party or his agent in "any suit or proceeding". The key requirement under s.8(2) is that the conduct sought to be admitted "influences or is influenced by any fact in issue or relevant fact". Logically, subsequent conduct that is used to prove a prior fact can only ever "be influenced" by the prior fact and so falls within this part of s.8(2). However, to be admissible, subsequent conduct must also be influenced by "any fact in issue or relevant fact". If we take for instance a case where the meaning of a contract is a fact in issue, there is no reason why subsequent conduct cannot fall within s.8(2) as being "influenced" by this fact in issue. Similarly, if the existence of a contract is a fact in issue, there is also no reason why subsequent conduct cannot be said to be "influenced" so as to be admissible by s.8(2). Indeed, these examples are analogous to the criminal cases where subsequent conduct is admitted freely as a matter of course: in those cases, the accused's conduct subsequent to the alleged offence is admitted by s.8(2) to prove the fact in issue, which is whether the alleged offence was committed. There is no reason why the outcome should be different just because the case concerns a contractual claim rather than a criminal offence. Accordingly, it is submitted that the EA does in fact allow subsequent conduct to be admitted under s.8(2), even for contract cases.

Common law restrictions incompatible with Evidence Act

If the EA does allow subsequent conduct to be admitted, what should we then make of the Singapore courts' pronouncements of either a complete prohibition

[104] Pinsler, *Evidence and the Litigation Process* (2010), p.171.
[105] *Lewis Christine v PP* [2001] 2 S.L.R.(R.) 131 at [29].
[106] *Krishna Jayaram v PP* [1989] 3 M.L.J. 272.
[107] *Chanderasekaran v PP* [1971] 1 M.L.J. 153.
[108] *Lewis Christine v PP* [2001] 2 S.L.R.(R.) 131.

or more restricted admission of subsequent conduct? In this regard, s.2(2) of the EA provides as follows:

> "All rules of evidence not contained in any written law, so far as such rules are inconsistent with any of the provisions of this Act, are repealed."

The effect of s.2(2) is thus to render ineffective all common law rules of evidence inconsistent with the provisions of the EA.

Accordingly, to the extent that the EA does not enact any restrictions on the admissibility of subsequent conduct (indeed, it allows for an expansive admission of such evidence by s.8(2)), the Singapore courts' pronouncements of a more restrictive use of subsequent conduct to determine a prior fact in contract law are inconsistent with s.8(2) of the EA and hence must be regarded as ineffective pursuant to the terms of s.2(2) of the same Act. There is a direct conflict between these rules of evidence, and the rule contained in the EA must prevail.[109]

More specifically, a similar argument particular to the use of subsequent conduct for contractual interpretation may be made by reference to ss.94 to 100 of the EA. The EA provides through these provisions specific instances where extrinsic evidence may (or may not) be admitted for the interpretation of a contract. These sections do not distinguish between subsequent conduct and other types of extrinsic evidence that can be used for the interpretation of contracts in Singapore. Accordingly, a common law development that justifies the *specific* exclusion of subsequent conduct would be inconsistent with these sections of the EA and hence must be ineffective.

It is therefore submitted that the Singapore courts' attempt to restrict the admissibility of subsequent conduct, where it is used to prove a prior fact in time under Singapore contract law, are ineffective for being inconsistent with the EA. This reason alone is enough to support the view that subsequent conduct should be admitted expansively under Singapore contract law. However, assuming for the sake of argument that it is wrong to conclude that the Singapore courts' restriction of the use of subsequent conduct is a rule of evidence as opposed to a rule of substance (which would in turn render the argument here inapplicable), let us now consider some independent policy reasons as to why subsequent conduct should be used expansively where it is used to prove a fact prior in time under Singapore contract.

Policy reasons against admission of subsequent conduct not convincing

Time of formation

Assuming, however, that the evidentiary point just made is wrong, it will now be argued that the policy reasons frequently advanced against the admission of subsequent conduct are not convincing and should be rejected. Many of these reasons have been convincingly refuted in the literature and so will only be dealt

[109] J. Pinsler, "Approaches to the Evidence Act: The Judicial Development of a Code" (2002) 14 S.Ac.L.J. 365, 375.

with briefly in this article.[110] Perhaps the most-cited argument is that it is impermissible to refer to subsequent conduct because such conduct, having arisen *after* the parties' intention had crystallised at the point of formation, cannot logically affect that intention in any way. For example, the Court of Appeal had stated in *Soup Restaurant* that "the object of the interpretative exercise was to discern the parties' intentions at the time of entering into the contract", and that reliance on such evidence is dangerous because "it can, with the benefit of hindsight, be shaped to suit each party's position".[111]

With respect, this is not a convincing argument. As Professor McLauchlan has pointed out, this argument confuses two logically distinct matters, i.e. the issue to be determined (the parties' intention at the time of contractual formation), and the admissible evidence (the parties' conduct after the time of contractual formation).[112] The admission of subsequent conduct does not mean that the parties' intention is no longer being assessed as at the time of formation. Subsequent conduct is simply being admitted as an aid to assessing the correct intention at the time of formation. Indeed, in relation to the use of subsequent conduct for contractual interpretation, Lord Nicholls has pointed out that[113]:

> "Evidence of the parties' subsequent conduct is sought to be used as a means of identifying the meaning which the language of the contract has borne from its inception. The fact that this evidence only came into being after the contract was made can hardly be a good reason for declining to admit it."

Moreover, as discussed earlier, the courts routinely admit subsequent conduct in criminal cases to assess whether an accused person is guilty of an offence. Obviously, the accused's guilt is determined at the time the alleged offence was committed. If the same argument that prohibits subsequent conduct for determining parties' intention prior in time is applied, the accused's subsequent conduct will not be admissible to prove his guilt. Yet, this is not the case. For example, it was held in *Krishna Jayaram v PP*[114] that an accused person's conduct of arrangement for a meeting with a person to change his statement to the prosecuting authority was admissible as relevant subsequent conduct. It is logically inconsistent to allow for subsequent conduct here, but to disallow subsequent conduct to determine parties' intention prior in time in the law of contract.

A slight variation of this policy against the admission of subsequent conduct is that such conduct is generally "unhelpful" towards assessing the parties' intention prior in time. However, "unhelpfulness" cannot be a relevant criterion to determine admissibility. At best, it should only be used towards the question of weight. Indeed, when courts consider whether subsequent conduct is "helpful", they have implicitly admitted such conduct already. Their decision not to base their findings on subsequent conduct is predicated not so much on the legal question of whether

[110] See, e.g., G. McMeel, "Prior Negotiations and Subsequent Conduct—the Next Step Forward for Contractual Interpretation?' (2003) 119 L.Q.R. 272; D.W. McLauchlan, "In Defence of a Role for Subsequent Conduct in Contract Interpretation" (2006) 12 N.Z.B.L.Q. 30; D.W. McLauchlan, "Contract Formation, Contract Interpretation, and Subsequent Conduct" (2006) 25 U.Q.L.J. 77; and D.W. McLauchlan, "Contract Interpretation and Subsequent Conduct" (2009) N.Z.L.J. 125.

[111] *Soup Restaurant* [2015] 5 S.L.R. 1187 at [74].

[112] McLauchlan, "In Defence of a Role for Subsequent Conduct in Contract Interpretation" (2006) 12 N.Z.B.L.Q. 30, 33.

[113] D. Nicholls, "My Kingdom for a House: The Meaning of Words" (2005) 121 L.Q.R. 577, 589.

[114] *Krishna Jayaram v PP* [1989] 3 M.L.J. 272.

the conduct is admissible, but rather on whether such conduct is of sufficient quality so as to affect the eventual decision reached.

Subjectivity

Another policy reason against the admissibility of subsequent conduct is that such conduct tends to reveal the parties' subjective intention. And because the law of contract is premised on objectivity, such intention can never be admitted. This reason was advanced by, for example, the High Court in *Re EG Tan & Co (Pte)*, where it appeared to equate the parties' subsequent conduct with their "subjective wishes or intent of either party", which it rightly held was eschewed in the interpretative process.[115] Two points may be made in response.

The more general point is that not all subsequent conduct will represent the subjective intention of the contracting parties. Thus, it would not be correct to lay down a blanket rule that prohibits all recourse to subsequent conduct on the premise that such conduct will be subjective in nature. To be fair, this has not been the approach taken by the Singapore courts. But even if the subsequent conduct turns out to be subjective, it is submitted that that characteristic alone should not be reason to deny its admission. This is because not all subjective belief will turn out to be unhelpful. Take for example the use of subsequent conduct in *Lewis Christine v PP*,[116] where the High Court ruled that an accused person's attempt to escape from custody has been admitted as a relevant fact to support the inference that she had commtitted theft. It is clear that whether an offence is made out is not based on the subjective belief of the accused person, but on an objective determination of law and fact. An accused person's escape is founded on his own subjective belief as to his guilt, and yet a criminal court, faced with the presumption of innocence in favour of the accused person, is nonetheless willing to take that subsequent conduct into account. There is no reason why a civil court in a contractual matter should not be able to do the same. As Professor McLauchlan says, such subsequent conduct "may be of sufficient weight to tip the scales in favour of a conclusion".[117]

The more specific point is in relation to the use of subjective intent to interpret a contract under Singapore evidence law. Unlike English law, provisions of the EA allow for the admission of subjective intent where the contract is latently ambiguous. Thus, as the High Court held in *HSBC Trustee (Singapore) Ltd v Lucky Realty Co Pte Ltd*, the EA mirrors the law of contract in allowing such evidence to be admitted where there is a larent ambiguity in the contract.[118] Thus, the Singapore courts' rejection of subsequent conduct on the basis that they represent the subjective intention of the parties, without more, may not be correct. Indeed, it may be said that the courts cannot do this, since the rule of evidence prescribed in the EA mandates a particular approach.

[115] *Re EG Tan & Co (Pte)* [1990] 2 S.L.R.(R.) 386 at [14]. See also *Jia Min Building Construction Pte Ltd v Ann Lee Pte Ltd* [2004] 3 S.L.R.(R.) 288 at [41].
[116] *Lewis Christine v PP* [2001] 2 S.L.R.(R.) 131.
[117] McLauchlan, "In Defence of a Role for Subsequent Conduct in Contract Interpretation" (2006) 12 N.Z.B.L.Q. 30, 37.
[118] *HSBC Trustee (Singapore) Ltd v Lucky Realty Co Pte Ltd* [2015] 3 S.L.R. 885 at [52].

Uncertainty

Yet another policy reason against the admission of subsequent conduct is that to do so would cause uncertainty. This view is clear from the Singapore courts' repeated citation of Lord Reid's speech in *James Miller*, where his Lordship spoke of the possibility that the meaning of a contract could be changed by reason of subsequent events.[119] Again, two points may be made in response. The first simply is that such evidence may, in appropriate cases, shed light on the parties' intention. As Lord Nicholls has said extrajudicially, prior negotiations (and by extension subsequent conduct) may "furnish a clear insight into the intended meaning of the disputed provision".[120] Indeed, as Professor McLauchlan has argued, Lord Reid's speech in *James Miller* must be read in the context of the case. Properly understood, the context clearly rendered the subsequent conduct in that case irrelevant. As such, Lord Reid's statement may well be confined to that particular instance only and not be regarded as a general point.[121]

A more specific point is that the current inconsistency in the admission of subsequent conduct under Singapore contract law has already introduced a degree of uncertainty. Thus the High Court has on various occasions pointed out the inconsistency between the liberal use of subsequent conduct for contractual formation and the more restrictive use of it in contractual interpretation respectively. Yet, the Singapore courts appear to tolerate the uncertainty occasioned by that inconsistency. If so, there little reason to object to the uncertainty supposedly caused by an expansive admission of subsequent conduct while preserving the consistency between the various uses of subsequent conduct.

Time and efficiency

A final policy reason frequently advanced against subsequent conduct is that an expansive consideration of it will waste time and result in inefficiency. The simple point in response simply is that interpretation claims, if proceeding on a contextual approach, will inevitably involve a wide consideration of the contextual evidence anyway.[122] Indeed, the whole basis of the contextual approach is the ascertainment of the parties' intention through a thorough investigation of the context. If there is an objection against the time and efficiency wasted by such an approach, the complaint should properly be directed against the contextual approach. However, if the contextual approach does apply, then it should be properly applied. This should result in the expansion admission of subsequent conduct to satisfy the reasons behind the contextual approach in the first place.

In summary, the policy reasons against the admission of subsequent conduct are not convincing. Indeed, the opposite is true: an adherence to the contextual approach requires that as much of the relevant context should be considered as possible. Any departure from this approach is to go against the rationale behind the contextual approach to ascertain the parties' intention in as true a fashion as

[119] *James Miller* [1970] A.C. 583 at 603.
[120] D. Nicholls, "My Kingdom for a Horse" (2005) 121 L.Q.R. 577, 589.
[121] McLauchlan, "In Defence of a Role for Subsequent Conduct in Contract Interpretation" (2006) 12 N.Z.B.L.Q. 30, 41.
[122] McLauchlan, "In Defence of a Role for Subsequent Conduct in Contract Interpretation" (2006) 12 N.Z.B.L.Q. 30, 45.

can be. The use of subsequent conduct is widely practised in other areas of law, such as in the area of criminal law; there is no reason why the courts should shy away from such evidence when it is being used for broadly speaking the same purpose in a civil case.

Conclusion

This article has first argued for a consistent approach in the use of subsequent conduct in Singapore contract law, where such evidence is being used to prove a fact prior in time. There is no reason why a different approach should be adopted with respect to using such conduct to interpret a contract, as compared with ascertaining the existence of a contract or the content of specific terms, when the ultimate fact being proved by the conduct is the parties' intention at the point of contractual formation. Indeed, this inconsistency is made all the more glaring when it is taken for granted that courts in criminal cases routinely consider the accused person's conduct after the commission of an offence to determine whether the accused was guilty in the first place. If it is accepted that the various uses of subsequent conduct to prove a fact prior in time should be consistent, this article has argued for an expansive use of subsequent conduct based on two reasons. The first is that the EA arguably mandates such an outcome, and any common law prescribing a more restrictive approach would be inconsistent with the EA and ineffective. However, even if it is regarded that the EA has no role to play, it is further argued that the policy reasons against the admission of subsequent conduct are not too convincing, especially if the contextual approach is being applied. The conclusion therefore is that, for reasons of both principle and policy, a consistently expansionist approach should be adopted towards the use of subsequent conduct under Singapore contract law to prove a fact prior in time. This would include the use of such conduct to interpret a contract, to ascertain the existence or nature of a contract, and to determine the specific terms of a partially oral contract.

Finally, while this article has focused on Singapore contract law, there is no reason why the argument premised on consistency cannot apply to other jurisdictions influenced by English law. While the evidential point is particular to those countries where the Indian Evidence Act, or its equivalent, apply, the need for consistency requires that subsequent conduct be used consistently in contract law influenced by the English common law, whether it is to determine contractual formation, or contractual interpretation.

Rethinking the Standard for Ascertaining the Quantum of Disgorgement in Patent Law Contexts

Moshood Abdussalam
University of Tasmania

keywords to be inserted by the indexer

Abstract

This article makes a case for the adoption of the "differential approach" over the "apportionment approach" as the proper criterion of determining the quantum of profits to be disgorged in patent enforcement. Its thesis is that the differential approach is more suitable to the peculiarities of the patent system simply because it accords with the utilitarian nature of the system. The novelty of this article lies in its suggestion that incremental benefits derived from patent infringement should be determined against the utilities or economic advantages enabled by the prior art before the infringed patent was issued.

Introduction

Disgorgement, otherwise known as an account of profits, is a legal remedy of equitable origin often deployed as an alternative to damages (i.e. compensatory damages and reasonable royalties) towards the enforcement of patents in most common law jurisdictions.[1] This is because, doctrinally, disgorgement and damages are considered to be mutually exclusive remedies in that the former aims to disgorge the unlawful gains derived by the infringer while the latter serves to compensate for losses suffered by the claimant.[2] Requiring the claimant to choose between these remedies helps to avoid dual reparation or recovery; it precludes the claimant from being compensated for losses and also being rewarded through a monetary measure of gains derived by the infringer.[3]

Disgorgement is an equitable remedy that treats the wrongdoer as bearing the duty of keeping all the gains derived from the infringing wrong for the proprietor of the entitlement as though the infringer were a (constructive) trustee.[4]

[1] Australian Patent Act 1990 (Cth) s.122; UK Patent Act 1977 s.61; see also S. Watterson, "An Account of Profits or Damages? The History of Orthodoxy" (2004) 24 *Oxford Journal of Legal Studies* 471.

[2] *Tang Man Sit (Deceased) v Capacious Investments Ltd* [1996] A.C. 514 (PC (Hong Kong); see also P. Birks, "Inconsistency between Compensation and Restitution" (1996) 112 L.Q.R. 375; see also J. Stevens, "Election between Alternative Remedies" (1996) 4 *Restitution Law Review* 117.

[3] *Tang Man Sit v Capacious Investments* [1996] A.C. 514; see also Birks, "Inconsistency between Compensation and Restitution" (1996) 112 L.Q.R. 375; see also Stevens, "Election between Alternative Remedies" (1996) 4 *Restitution Law Review* 117.

[4] Per Lloyd-Jacob J in *Electrolux v Electrix Ltd* (1953) 70 R.P.C 158 at 159: "The principle upon which the court grants an account of profits, as I have always understood is to be, is this, that where one party owes a duty to another,

Notwithstanding the term "trustee", the infringer's mandate to account for profits derived from the wrongful act is personal, not proprietary.[5] The infringer is accordingly treated as a debtor, whose state of insolvency does not vest the claimant with the status of a secured creditor over the assets of the infringer.[67]

As with other equitable remedies, disgorgement is discretionary, resting on equitable considerations.[8] Factors such as undue delay, laches and acquiescence, and knowledge borne by the wrongdoer have bearing upon how and whether the remedy would be awarded,[9] albeit subject to the terms of any applicable statute.[10]

Disgorgement is a remedial strategy that ostensibly aligns more closely to the utilitarian nature of patent law than does damages.[11] The utilitarian nature of patent law lies essentially in the objective of the system to reward inventors and their sponsors for the incremental value of the inventions.[12] While the superiority of disgorgement over damages, from this utilitarian perspective, falls outside the scope of this article, it does address how the standard for disgorgement can be devised to reflect this utilitarian foundation.

It makes a case for the adoption of the "differential approach" over the "apportionment approach" as the proper criterion for determining the quantum of profits to be disgorged in patent enforcement. This was in turn prompted by the lack of uniformity among major common law jurisdictions in the ascertainment of the quantum of disgorgement in patent law context.

The article argues that the differential or incremental approach provides an optimal standard for the ascertainment the quantum of disgorgement, a thesis founded on two premises: (1) that the definition of "profits" should be perceived from an economic or holistic standpoint rather than an accounting-based one; and (2) that fairest measure of "profits" is best ascertained on the basis of the incremental benefits accruing to the infringer as facilitated by incremental technical advantage furnished by the infringed patent. Its novelty lies in its suggestion of an improvement to the differential or incremental approach with the introduction of a significant element: gauging the incremental benefits derived from a patent infringement against the utilities or economic advantages enabled by the prior art before the infringed patent was issued.

The disgorgement remedy generally has two components: the quantum of disgorgement and the deduction of expenses. The article targets the second of these, which has its own elements, namely the definition of the term "profits" and

the person to whom that duty is owned is entitled to recover from the other party every benefit which that other party has received by virtue of his fiduciary position …").

[5] See D. Sheehan, "Subtractive and Wrongful Enrichment: Identifying Gain in the Law of Enrichment" in C. Rickett (ed.), *Justifying Private Law* (Oxford: Hart Publishing, 2008), p.531; see also D. Ong, "Breach of Fiduciary Duty: The Alternative Remedies" (1999) 11 *Bond Law Review* 336.

[6] See K. Barnett, "Disgorgement of Profits in Australian Private Law" in E. Hondius and A. Janssen (eds), *Disgorgement of Profits: Gain-Based Remedies throughout the World* (Switzerland: Springer International, 2015), p.19.

[7] Barnett, "*Disgorgement of Profits* in Australian Private Law" in Disgorgement of Profits (2015), p.19.

[8] *Novoship (UK) Ltd v Mikhaylyuk* [2015] Q.B. 499 CA (Civ Div) at 535 ("We consider that where a claim for an account of profits is made against one who is not a fiduciary, and does not owe fiduciary duties then, as Lord Nicholls said in the *Blake* case [2001] 1 AC 268, the court has a discretion to grant or withhold the remedy").

[9] F. Patfield, "The Modern Remedy of Account" (1987) 11 *Adelaide Law Review* 1.

[10] See, for example, UK Patent Act (UK) s.62 (which excludes an account of profits in circumstance where the infringer lacked knowledge, at the date of infringement, about the existence of the enforced patent).

[11] D. Opderbeck, "Patent Damages Reform and the Shape of Patent Law" (2009) 89 *Boston University Law Review* 137.

[12] J. Schlicher, "Patent Damages, the Patent Reform Act, and Better Alternatives for the Courts and Congress" (2009) 91 *Journal of The Patent and Trademark Office Society* 19.

the ascertainment of the quantum of disgorgement (or causation). In addressing these elements with a view to pursuing the above thesis, this article comprises four major sections. The first section addresses the term "profits", the definition of which is judicially disputed and lacking in uniformity. The second section introduces the quantification of the quantum of disgorgement and the pith of competing approaches for its ascertainment—i.e. the apportionment and the differential or incremental approaches. The third section critically discusses the features and practical implications of these competing approaches. Finally, the fourth section makes a case for the differential approach in patent law contexts.

It is also important to state up front that this article draws largely from the common law jurisprudence and academic literature of Commonwealth nations. While it also draws on relevant case law from US jurisprudence and academic literature on the disgorgement remedy to patents, the body of US case law brought to bear is essentially old cases applying the remedy to patents, antedating the US Patent Act of 1948, which abolished the application of the remedy to patents. Yet there surface calls in the US for resuscitating the remedy[13] stemming from an outcry against "patent holdup" problems (i.e. disproportionality between the incremental value of patents and remedial outcomes) facilitated by damages.

Definition of profits

The lack of uniformity in the definition of the term "profits" for the purposes of the disgorgement remedy is first addressed. Although the basic purpose of the disgorgement remedy is to deprive or strip an infringer of the gains derived from infringing activities, the proper definition of the term "profits" is not without difficulty and significant controversy. In the seminal Australian High Court case of *Colbeam Palmer v Stock Affiliates*,[14] Windeyer J, acknowledging the difficulty in defining the term, said:

> "In modern economic theory the profit of an enterprise is a debatable concept. Consequently, the word 'profit' has today varying senses in the vocabulary of economists."[15]

Despite this acknowledgement, he considered that it was only right for him to adopt the definition of the term in consonance with judicial precedent.[16] Thus his Honour followed the definition expressed by Lord Lindley in *Re Armitage*,[17] which defined profits as the difference between what goes into an undertaking, and what comes out of it. This definition is well exemplified in the case of *Unilin Beeher BV v Huili Building Materials Pty*[18] where Allsop J said, among other things, that:

> "The infringer should account for the actual profits that they have gained from the infringement. In calculating those profits, it is appropriate to deduct

[13] See A. Frye, "'Inextricably Commingled': A Restitution Perspective in Patent Remedies" (2013) 26 *Harvard Journal of Law and Technology* 669; see also C. Roberts, "The Case for Restitution and Unjust Enrichment Remedies in Patent Law" (2010) 14 *Lewis and Clark Law Review* 101.
[14] *Colbeam Palmer v Stock Affiliates* (1968) 122 C.L.R. 25.
[15] *Colbeam Palmer* (1968) 122 C.L.R. 25 at 37.
[16] *Colbeam Palmer* (1968) 122 C.L.R. 25 at 37.
[17] *Re Armitage* (1893) 3 Ch. 337 CA.
[18] *Unilin Beeher BV v Huili Building Materials Pty (No.2)* [2007] FCA 1615.

> costs directly attributable to selling and delivering infringing articles from the revenue made from such sales and deliveries."[19]

It is important at this juncture to explore the two different definitions of the term "profits" for the purposes of this remedy. These are the accounting-based and the economic or holistic definitions of profits. It is significant to reiterate that the economic or holistic definition of profits is superior to the other definition, and why this is so will be justified in the course of the discussion on the subject.

Accounting-based definition of profits

This viewpoint on the definition of profits is informed by the tradition of practitioners of the accountancy profession. In accounting practice, profit is considered as the difference between the costs of production and the total revenue gained from an activity.[20] Apparently, such a definition excludes from the conception of "profits" other economic advantages gained by the infringer which are non-monetary. There appear to be three major lines of argument upon which this accounting-based conception is adopted in the application of the disgorgement remedy. The first is that the infringer is considered the patentee's agent or constructive trustee who is expected to act efficiently and resourcefully, which could consist of reducing production costs. The reason for this, as explained by Laddie J in *Celanese International Corp v BP Chemicals*, is that the infringer "is treated as if he conducted his business and made the profits on behalf of the plaintiff".[21]

The consequence of this reasoning is that an infringer is not to be considered to have made profits until he or she has positively earned, in monetary terms, from their infringing activity. Going by this reasoning, as reflected in the words of Laddie J in *Celanese International*:

> "If an infringer's process makes no profit overall, then whether the infringement accounts for 10 per cent or 1000 per cent of the profits … the plaintiff will recover nothing."[22]

He goes on to say that "[a] large percentage of zero is still zero".[23] Storrow, writing in 1878, reasoned along the same lines as Laddie J, saying:

> "'Profit' is the gain made upon any business or investment when both receipts and payments are taken into account. And yet the so-called rule of savings contradicts this; for it looks only into a comparison of the different payments and costs with each other, and not with receipts. Whosoever forgets that there is no 'profit' in any sense until there is an actual receipt of money or a saleable product of an ascertained market-value to set off against cost, will come in practical affairs to insolvency, and in law to an absurdity."[24]

[19] *Unilin Beeher* [2007] FCA 1615 at [70].
[20] C. S. Bradford, *Basic Accounting Principles for Lawyers*, 3rd edn (Lexis Nexis, 2014), pp.11–13.
[21] *Celanese International Corp v BP Chemicals* [1999] R.P.C. 203 Ch D at 219.
[22] *Celanese International* [1999] R.P.C. 203 at 217.
[23] *Celanese International* [1999] R.P.C. 203 at 217.
[24] J. Storrow, "Money Recoveries in Patent Suits" (1879) 13 *American Law Review* 7.

Edelman identifies the second line of argument on which the accounting-based conception of profits stands to be the belief of courts that the law is not interested in depriving persons of the opportunity of availing themselves of cost-reducing avenues if other legitimate avenues would have caused them loss.[25] Further to this, Edelman identifies the third reason for the definition to be that savings in expenses derived from wrongdoing are considered to be outside the precinct of the remedy.[26] This third reason can be fairly described as weak in the patent context because the judicial authority upon which Edelman founds his argument is *Attorney General v Blake*,[27] a breach of contract case where disgorgement was applied.

Economic or holistic definition of profits

It is has been argued that the accounting-based conception of profits is very shallow and ignores the economic nature of the patent system with regard to the utility of inventions. Roach posits that a judicial assessor concerned with a determination of "profits" derived from the unauthorised application of a patent, which is a utilitarian entitlement, should first direct focus to the overall economic advantage derived from the infringement of an IP right, then move on to ascertaining the monetary equivalent of that advantage.[28] In other words, the benefits derived from a patent could be more than merely the immediate monetary gains.[29] Benefits derived from infringement could include savings or reductions in the costs of production. It could also be in the form of an advantage that helps the infringer to overcome barriers or hurdles (e.g. regulatory requirements or learning-curve) that would naturally have delayed its entry into the market upon the expiration of the patent.

The British textbook writers Blanco White et al. also assert that the value of savings needs to be taken into account for the purposes of determining account of profits.[30] They hypothesise a situation where a particular patented method for making stocking toes "saves on average, per dozen pairs of stockings, so many minutes of operative time at so much an hour".[31] They reason that when the patentee decides to make an election between compensatory damages and the disgorgement remedy, the patentee would have had to consider the value of the savings to the infringer before making such an election between both alternative remedies.[32]

Cotter considers it counterintuitive that non-monetary gains derived from the unauthorised application of patents would not be treated as profits,[33] particularly as certain inventions are directed at reducing production costs, and the infringement of certain patents is done with a view to gaining particular economic advantages.

[25] J. Edelman, *Gain-Based Damages: Contract, Tort, Equity and Intellectual Property* (Oxford: Hart Publishing, 2000), pp.74–75.

[26] Edelman, *Gain-Based Damages* (2000), pp.74–75.

[27] *Attorney General v Blake* [2001] 1 A.C. 268 HL at 291.

[28] G. Roach, "Counter-Restitution for Monetary Remedies in Equity" (2011) 68 *Washington and Lee Law Review* 1271, 1281–1286.

[29] See W.M. Macomber, "Damages and Profits in Patent Cases" (1910) *Columbia Law Review* 639; see also H.A. Toulmin, "Problems in Profits and Damages in Patent Accounting" (1915) 2 *Virginia Law Review* 507.

[30] B. White et al., *Patents, Trade Marks, Copyright and Industrial Designs*, 2nd edn (London: Sweet and Maxwell, 1978), pp.9–10.

[31] White et al., Patents, *Trade Marks, Copyright and Industrial Designs* (1978), pp.9–10.

[32] White et al., *Patents, Trade Marks, Copyright and Industrial Designs* (1978), pp.9–10.

[33] T. Cotter, *Comparative Patent Remedies* (Oxford: Oxford University Press, 2013), p.201.

A good example is *Siddell v Vickers*.[34] In this case, the patent in issue was a mechanical appliance for turning large ingots in the course of forging iron or steel. The patented invention was an improvement over manual labour that had been previously adopted by the infringer in the steel forging process, and its merits lay in its cost-saving facility. The court made a disgorgement award on the basis of the savings attained by the infringer, not on the basis of the steel produced from the infringing application of the machine. This is because the steel itself was neither patented nor novel in nature.[35] This position is, however, controverted by Laddie in *Celanese International*, as discussed below.

Also, in *Bayer Corpscience v Charles River Lab*,[36] the infringer was able to satisfy regulatory requirements and wait until the patent expired before swiftly entering the market. This helped the infringer avoid market entry delays that it would have encountered in trying to meet regulatory requirements. Lord Malcolm awarded disgorgement on the basis of the gain derived from the swift market entry enabled by the infringement, ruling that all that mattered was whether "there is a sufficient link or nexus between the wrong and the ultimate financial consequences".[37]

The implication of an economic or holistic definition of profits can be cumulative, such that the infringer can be required to disgorge not only the immediate monetary gains (in accounting terms) derived from the infringement but also the monetary value of savings in the costs of production. This can be illustrated using a simple hypothetical example. Suppose, prior to infringement, a scooter manufacturer produced a model that required the employment of 15 workers to produce 100 pieces weekly. Let it be assumed that $250 was earned for each scooter of that model sold. However, by switching to an infringing model, scooters with better technical features and market appeal were produced, which the infringer sold at $300 per piece. Also, the manufacturer now required 10 workers to reach its weekly production target of 100 pieces. If the economic or holistic definition of profits were to be applied, the infringer will account not only for the measure of profits attributable to the infringement—the outcome of which will depend on whether an apportionment or incremental approach applied to assessing the quantum of disgorgement. The infringer will also account for savings in workers' salaries or wages that the infringing model enabled.

The need to define "profits" from an economic or holistic standpoint

It is submitted that courts should treat as "profits" all direct gains or beneficial changes, whether they be monetary gains or improvements in the economic conditions of the infringer, so long as they would not have been attained without the infringement. As rightly stated in the US case of *Schnadig Corp v Gains Manufacturing Co*,[38] the ultimate purpose of the disgorgement remedy is to "recover every dollar of advantage realized by the infringer from the infringement and no

[34] *Siddell v Vickers* [1892] R.P.C 152; see also F. Gilgo, *The Foundations of Restitution for Wrongs* (Oxford: Hart Publishing, 2007), pp.75–76.

[35] F. Patfield, "Remedy of Account of Profits in Industrial and Intellectual Property Litigation" (1984) (Special Issue) *University of New South Wales Law Journal* 189, 205.

[36] *Bayer Corpscience v Charles River Lab* [2010] CSOH 158.

[37] *Bayer Corpscience* [2010] CSOH 158 at [8].

[38] *Schnadig Corp v Gains Manufacturing Co* 206 U.S.P.Q. 202 (1980).

more".[39] It has been reasoned that there can be no hard and fast rules as to defining profits because different factual situations will bear different considerations of what amounts to profits.[40] However, as is shown in this article, the accounting-based definition of profits is informed by the approach to quantification of profits chosen by the courts—i.e. the apportionment approach. The economic or holistic view of profits, which is better, and which is advocated in this article, is enabled by a different approach to determining causation—i.e. the differential or incremental approach.

A robust view on profits founded on the economic realities of the occasion accords with the purposes of patent law. Patent law is aimed at enabling inventors to capture a measure of the social value they have contributed by virtue of the utility or technical advantage their patented inventions provide. The inventive utility or technical advantage disclosed in the patent specification and declared in the patent claim(s) is the focal point of protection in patent law.[41] A patent right represents what the patentee asserts that his or her invention can do and is the foundation of protection under patent law.[42] Therefore, if reductions in the costs of production (e.g. resources such as time, materials or labour) are the technical advantages derivable from the application of an invention, this will contradict the purpose of the patent system not to found an assessment of disgorgement on the basis of these technical advantages.

To circumscribe the definition of profits to an accounting-based one has the propensity to place patentees at a distinct economic disadvantage, so that in court decisions and in private settlements reached on the basis of the remedy, the infringer is likely to have a favourable edge over patentees. The full extent of the economic advantages derived from infringement would not be accounted for, especially where the inventions enable advantages that are not immediately translated into money, such as where they reduce production costs or bestow other (secondary) benefits. The implication of this is that infringers might feel emboldened to infringe. Where this is the case, inventors and their sponsors might have reduced capabilities to recoup their marginal costs.

Having addressed the term "profits" and the debate that surrounds its proper definition, it becomes necessary to move focus towards the "quantum of disgorgement" element of the remedy.

Causation and quantum of disgorgement

While defining profits is an essential component of the disgorgement remedy, there is a quantification exercise upon which the substance of the remedy's application pivots, which entails determining the measure of "profits" attributable to the infringement.[43] This exercise is subject to the rules of "causation"—causal connections between the profits earned and the legal wrong done.[44] The High Court

[39] *Schnadig Corp* 206 U.S.P.Q. 202 (1980).
[40] *Levin Bros v Davis Mfg Co* 72 F. 2d 163 (8th Cir. 1934).
[41] See, for example, *Merck & Co Inc v Generics UK (Ltd)* [2004] R.P.C. 31 Ch D at [38].
[42] See C. Bodkin, *Patent Law in Australia*, 2nd edn (London: Thomas Reuters, 2008), pp.375–381.
[43] See *Zupanovich v B and N Beale Nominees Pty* [1995] FCA 1424; see also, *McCambridge Ltd v Joseph Brennan Bakeries* [2014] IEHC 269.
[44] See T. Moody-Stuart, "Quantum in Accounts of Profits: the Acid Test" (1999) 21 E.I.P.R. 147; see also L. Bently, "Accounting for Profits Gained by Infringement of Copyright: When Does it End?" (1991) 13 E.I.P.R. 5.

of Australia, in *Warman International v Dwyer*,[45] reasoned that a more probative approach in patent and other IP contexts is required simply because in these contexts there is greater concern to avoid unjust enrichment from either the right owner or the infringer.[46] In other words, this approach prevents any opportunistic advantage from either side that is not intended by the law. This position was confirmed by the UK Court of Appeal in *OOO Abbott v Design and Display Ltd*,[47] where the court said, among other things, that "the purpose of the account is to quantify the extent to which the infringer will have been unjustly enriched".[48] One can only assume, rightly, that the utilitarian, but not libertarian, foundation of IP rights is the reason for such caution.

The quantification exercise entails determining the causal connection between the infringement and the gains derived from the infringement, with a view to ascertaining the measure of disgorgement to be made to the patentee. As correctly expressed obiter in *The United Horse-Shoe*, whenever the disgorgement remedy is elected, it becomes necessary in ascertaining the profits to be disgorged, to "determine what proportion of the net profits realised by the infringer was attributable to its use".[49]

The perceived difficulty of the disgorgement remedy

The quantification exercise is considered to be problematic and difficult for both judges and parties concerned in the exercise.[50] The reason for this, as explained by DeMott, is over-determination[51]—a situation where a multiplicity of (market) factors is responsible for the wrongdoer's profits beyond the wrongful breach of the claimant's entitlement.[52] As DeMott puts it "[o]ver-determination can compound quantification problems when the fiduciary's own efforts legitimately contributed to the profit".[53] In *Siddell v Vickers*,[54] Kekewich J made a statement that over time has become a popular description of the nature of difficulties that assail the quantification exercise. He said:

> "I do not know any form of account which is more difficult to work out, or may be more difficult to work out than an account of profits … and the difficulty of finding out how much profit is attributable to any one source is extremely great—so great that accounts in that form very seldom result in anything satisfactory to anybody. The litigation is enormous, the expense is great, and the time consumed is out of all proportion to the advantage ultimately attained … ."[55]

[45] *Warman International v Dwyer* (1995) 182 C.L.R. 544.
[46] *Warman International v Dwyer* (1995) 182 C.L.R. 544 at [23]–[33].
[47] *OOO Abbott v Design and Display Ltd* [2016] EWCA Civ 98; [2016] F.S.R. 27.
[48] *Design and Display Ltd v OOO Abbott* [2016] EWCA Civ 98; [2016] F.S.R. 27 at [7].
[49] *United Horse-Shoe and Nail Co Ltd v John Stewart & Co* [1888] R.P.C. 266 HL.
[50] See G. Dike, "Trial of Patent Accountings in Open Court" (1922) 36 *Harvard Law Review* 33.
[51] D. DeMott, "Causation in the Fiduciary Realm" (2011) 91 *Boston University Law Review* 851, 858; see also M. Gergen, "Causation in Disgorgement" (2012) 92 *Boston University Law Review* 827.
[52] For an academic analysis of "overdetermination", see R. Wright, "Causation in Tort Law" (1985) 73 *California Law Review* 1735, 1775.
[53] DeMott, "Causation in the Fiduciary Realm" (2011) 91 *Boston University Law Review* 851, 858.
[54] *Siddell v Vickers* [1892] R.P.C. 152.
[55] *Siddell v Vickers* [1892] R.P.C. 152 at 162–163.

Approaches to determining causation and quantum

There are two broad ways of determining causation and measuring the quantum of profits to be disgorged: these are the apportionment and incremental (or differential) approaches.[56] The apportionment approach involves apportioning profits among the different elements or factors responsible for profits earned by the infringer.[57] In other words, the approach tries to determine the extent to which the infringed invention accounts for the profits earned by the infringer among other contributing non-infringing factors. In following this approach, where it appears that the entire market value of the infringing matter is derived from the infringed patent, the patentee will be entitled to the whole profits gained from the infringing undertaking without apportionment.[58] However, where there are other contributing causes of profits, the profit earned from the infringement will be apportioned among these other factors on the basis of their contributing values.

On the other hand, the incremental or differential approach (also known as the comparative approach) compares what the infringer made from the infringement with what the infringer could have made if he/she had applied non-infringing alternatives (NIAs) instead of infringing.[59] In other words, it compares what the infringer could have made applying their best available non-infringing alternative with what the infringer made from the incorporation of the infringed the patent.[60] The apportionment approach is generally used in Commonwealth jurisdictions, except in Canada where the differential or incremental approach is becoming increasingly popular.

At this juncture, a detailed comparison of both competing approaches is necessary in order to show the differences in legal nature.

Apportionment

As can be discerned from the US Supreme Court in *Garretson v Clark*,[61] where the patent does not create an entirely new product but an improvement, or where it is used in combination with other matters to produce an outcome, apportionment becomes necessary. The patentee "must separate its result distinctly from those of other parts, so that the benefits derived from it may be distinctly seen and appreciated".[62] In *Imperial Oil v Lubrizol*,[63] where the infringer motor oil business had added the patented and infringing additive to its brand of oil products, the Canadian Court of Appeal made an apportionment of profits between the motor oil product (without the patented additive) and the value attributable to the additive (as added to the motor oil). The court reasoned that:

[56] *Celanese International* [1999] R.P.C. 203 at 212–218; see also C. Colston and J. Galloway, *Modern Intellectual Property Law*, 3rd edn (Abingdon: Routledge, 2010), pp.789–790.

[57] *Cartier v Carlile* (1862) 31 Beav. 292 at 298.

[58] *Westinghouse Elec. Co v Wagner Elec. Co* 225 U.S. 604 (1912): "Where the patent, though using old elements, gives the entire value to the combination, the plaintiff is entitled to recover all of the profits"; see also *Hurlbut v Schillinger* 130 U. S. 472 (1889).

[59] P. Devonshire, *Account of Profits* (London: Thomson Reuters, 2013), p.138.

[60] See N. Siebrasse, "A Remedial Benefit-Based Approach to the Innocent-User Problem in the Patenting of Higher Life Forms" (2004) 20 *Canadian Intellectual Property Review* 79.

[61] *Garretson v Clark* 111 U.S. 120 (1884).

[62] *Garretson v Clark* 111 U.S. 120 (1884).

[63] *Imperial Oil v Lubrizol* [1997] 2 F.C. 3.

> "While motor oil containing the dispersant additive was properly claimed in the patent (it would seem likely that the dispersant is useless except as an additive to motor oil) and while that claim was properly found to have been infringed, the reality is that Lubrizol did not invent motor oil and that Imperial's motor oils contain other additives than the one here in issue."[64]

Similarly, in another Canadian case, *Wellcome Foundation Ltd and Glaxo Welcome v Apotex*,[65] apportionment was applied. In this case the infringer (Apotex) had applied the patented active ingredient known as TMP to its drug product, which also contained another non-infringing, active ingredient known as SMX. The court considered that the infringed TPM accounted for 60 per cent of the profits earned, while the remaining 40 per cent was ascribed to the non-infringing SMX.

In exceptional cases, where the entire profits earned from the infringement are ascribable to the patent such that they would not have been earned if the patented invention had not been applied by the infringer, then there will be no need for apportionment.[66] This principle was laid down in *Peter Pan Manufacturing Corp v Corsets Silhouette Ltd*,[67] a UK trade secret case, where the court ruled that the entire profits earned could not have been earned if the plaintiff's right had not been applied to the infringing products. This reasoning was equally applied in *Unilin Beeher v Huili*,[68] where the patent had become an industrial standard. It was held that the infringement of the patent was the reason the infringer earned the profits because, without infringement, the infringer's product could not have complied with the requirements of standardisation in the industry in question.[69]

However, the burden of proving apportionment is generally placed upon the infringer. Thus in *Ductmate Industries Inc v Exanno Products Ltd*[70] the court held that because the infringer had refused to show reason why an apportionment must be made, especially by refusing to provide details about its business, apportionment had to be refused.[71]

Differential/incremental value approach

In case of *Percy Schmeiser and Schmeiser Enterprises Ltd v Monsanto Canada Inc and Monsanto Co*,[72] which marks the Canadian shift towards the differential/incremental approach, the Canadian Supreme Court described the approach by saying:

> "The preferred means of calculating an accounting of profits is what has been termed the value-based or 'differential profit' approach, where profits are allocated according to the value contributed to the defendant's wares by the patent … A comparison is to be made between the defendant's profit

[64] *Imperial Oil v Lubrizol* [1997] 2 F.C. 3.
[65] *Wellcome Foundation Ltd and Glaxo Wellcome v Apotex* [2001] 2 F.C.R. 618.
[66] *Colbeam Palmer Ltd v Stock Affiliates Pty Ltd* (1968) 122 C.L.R. 25 at [9].
[67] *Peter Pan Manufacturing Corp v Corsets Silhouette Ltd* [1964] 1 W.L.R. 96 Ch D.
[68] *Unilin Beeher v Huili* [2007] FCA 1615.
[69] *Unilin Beeher v Huili* [2007] FCA 1615 at [80] ("The evidence discloses that the panels were described as the 'industry standard' in the October 2004 issue of a publication called *Flooring Australia and New Zealand*").
[70] *Ductmate Industries Inc v Exanno Products Ltd* 1987 C.L.B. 4452; see also *Reading & Bates Construction Co v Baker Energy Resources Corp* [1995] 1 F.C. 483.
[71] *Ductmate Industries v Exanno Products* 1987 C.L.B. 4452 at [12]–[13].
[72] *Percy Schmeiser and Schmeiser Enterprises v Monsanto Canada Inc and Monsanto Co* [2004] 1 S.C.R. 902.

> attributable to the invention and his profit had he used the best non-infringing option."[73]

The court held that, upon a comparison of what the infringer gained from infringing the patent and what it could have gained without it, no profit was earned from the infringement and for this reason the patentees were disentitled to an account of profits.[74]

This approach looks at the infringer's opportunity cost or alternative forgone in determining the measure of gain derived by an infringer from the infringement.[75] As stated by Zinn J in *Monsanto Canada v Rivett*,[76] the infringer's next best non-infringing alternative serves "as a baseline from which to calculate added value".[77] According to Zinn J this alternative does not need to be physically available but needs only to exist and be feasible.[78] Accordingly, if a consideration that the infringer could have adopted that alternative were ignored in the computation of profits to be disgorged, then the patentee would be able to receive the infringer's profits based on values (i.e. integers) that do not arise from the patented invention.

Furthermore, Laddie J in *Celanese International Corp*, in trying to analyse how the differential approach works, made two points: (1) that apportionment and the differential/incremental approach are mutually exclusive, and (2) that both competing approaches perceive profits from different viewpoints along the accounting/economic dichotomy of profits definition. As regards the first point of distinction, that the differential approach excludes any sort of apportionment, he said:

> "In the incremental approach it is neither necessary nor relevant how many steps or integers there are in the process or article nor is it relevant what each one contributes, if anything, to overall profitability. Under the incremental approach one is only looking at whether the infringing step is financially advantageous to the defendant when compared to the most likely alternative."[79]

Before saying this, however, he identified the second basis for distinction. This is that the conception of profit by the two approaches can be different. According to his Honour, as regards apportionment and its conception of profits,

> "if an infringer's process makes no profits overall, then whether infringement accounts for 10 per cent or 100 per cent of the profits, on an apportionment the plaintiff will recover nothing".[80]

In other words, if the infringer does not appear to have made profits from the infringement but is able to have reduced production by dint of the infringement, it cannot be considered that the infringers have made profits. However, as regards

[73] *Schmeiser v Monsanto Canada* [2004] 1 S.C.R. 902 at [102].
[74] *Schmeiser v Monsanto Canada* [2004] 1 S.C.R. 902 at [102].
[75] *Reading & Bates Construction v Baker Energy Resources* [1995] 1 F.C. 483; see also *Bristol-Myers Squibb Co v Canada (Attorney General)* [2005] 1 R.C.S. 533.
[76] *Monsanto Canada v Rivett* [2010] 2 F.C.R. 93; see also *Monsanto Canada Inc v Rivett and Monsanto Canada Inc v Janssens* [2010] FCA 207.
[77] *Monsanto Canada v Rivett* [2010] 2 F.C.R. 93 at [56].
[78] *Monsanto Canada v Rivett* [2010] 2 F.C.R. 93 at [63].
[79] *Celanese International* [1999] R.P.C 203 at [34].
[80] *Celanese International* [1999] R.P.C 203 at [31].

the differential approach and its conception of profit, Laddie J stated that "if a process makes a loss of £1m p.a. with the infringing step but would have made a loss of £3m p.a. without it, the benefit to the infringer is £2m p.a.".[81]

While these analyses by his Honour appear correct, there is a problem with how Laddie J chose to advance the second point in his attempt to show that the differential approach is alien to the Commonwealth. To advance this position, Laddie J relied on *Siddell v Vickers*. He treated this obviously differential approach case as one of apportionment. He reasoned that the Court of Appeal in that case had considered the entire market value of the infringing products as being attributable to the infringed patent. However, a close reading of the facts of the case shows that the infringed patent did not improve the marketability of the infringing products. It only reduced the cost of production, which made the infringer earn more than he would have otherwise earned had he adopted the manual production process previously used.[82]

In fact, both the Master and the presiding judge, Kekewich J, had disagreed over the preferred line of comparison of alternatives that ought to be taken to compute profits. The Master considered that the profits to be disgorged should be based on the comparison of the infringing method and another mechanical option later available to the infringer; the resulting difference being £235. On the other hand, Kekewich J preferred a comparison of the gains derived from the use of the infringing machine and the manual process that the infringer had employed prior to infringement, the result of which was £2,016. However, Laddie J acknowledged the likelihood that a differential approach was used when he said:

> "In my view it is impossible to read this decision as clearly indicating that an incremental approach … is correct. But if it was, it is not clear that the issue was before the Court of Appeal for determination since some sort of comparative assessment was being argued for by both parties."[83]

The central reason why Laddie J considered that the differential approach was not applied was because the Court of Appeal ordered the infringer to pay £3,000, which is £984 in excess of the sum Kekewich J had ordered using a differential approach. But this reason in itself does not justify Laddie J's view that apportionment governed the £3,000 award. This is so because in reaching that sum the court compared the infringer's NIAs with the infringed patent.

Another equally significant case, although neither a patent case nor one bordering on cost savings, on which Laddie J founded his reasoning that the differential approach is alien to the Commonwealth, is *My Kinda Town*. In this trade mark case, Slade J clearly adopted a differential approach to determine profits to be disgorged on the basis of comparison between what was earned using the infringing mark and what would have been earned without it.[84] However, Laddie J, in his attempt to disprove the argument that the differential approach was applied by Slade J in *My Kinda Town*, unconvincingly described the comparative exercise in

[81] *Celanese International* [1999] R.P.C 203 at [31].

[82] Per Muldoon J in *Lubrizol Corp v Imperial Oil* 1996 C.L.B. 7481: "In *Siddell v. Vickers*, as opposed to the present case, there was no price directly attributable to the turning. The turning was just a part of the manufacturing process, which part was not subcontracted out and therefore not separately priced. The difference between the profit without infringing and the profit when infringing was the best measure of all of the profit derived from infringing."

[83] *Celanese International* [1999] R.P.C. 203 at [71].

[84] *My Kinda Town Ltd (t/a Chicago Pizza Pie Factory) v Soll* [1982] F.S.R. 147 Ch D at 158.

that case as meant only to "discover what proportion of the defendant's customers were confused".[85] Another confusing feature is that Slade J inadvertently described *Siddell v Vickers* as an apportionment case when in fact what he meant in substance was that disgorgement tries to ensure that the claimant's reward is commensurate with the gains derived by the infringer from the infringement.[86]

Encapsulation

As rightly identified by Laddie J, the two approaches can lead to almost the same result in certain circumstances, but this will be largely coincidental because in most cases they are most likely to result in different outcomes.[87] However, contrary to this, Friedman notes that apportionment and differentiation are essentially the same as they are aimed simply at determining what quantum of profits should be disgorged to the patentee.[88] Friedman reasons that an apportionment could also be pursued by a differential approach. He argues that in *Metallic Rubber Tire Co v Hartford Rubber Work*[89] and *P.P Mast and Co v Superior Drill Co*[90] apportionment was attained by differential comparison. The truth is that Friedman's position is untenable as both approaches are most certainly likely to produce different outcomes than have similar ones. To show how applying either approach can bring about different outcomes, the next section explores the likely features of applying either of the approaches.

Likely features of applying the apportionment and differential approaches

Having discussed the characteristics of the two different approaches to determining the quantum of profits to be disgorged, it becomes necessary at this juncture to separately address the likely features of applying the each of the competing approaches.

The likely features of apportionment

There are two prominent features of apportionment: the concept of base allocated profits and uncertainties in the parameters of apportionment. These features will now be expounded upon.

Base allocated profits

One essential feature of the apportionment approach, as identified by Laddie J, is the concept of "base allocated profits".[91] According to his Honour, where profit is apportioned between the various elements incorporated in an infringing matter, as between infringing elements and non-infringing elements, the minimum value or

[85] *Celanese International* [1999] R.P.C. 203 at [65].
[86] *My Kinda Town* [1982] F.S.R. 147 at 155.
[87] *Celanese International* [1999] R.P.C. 203 at [31] ("If these two routes arrive at similar figures it is a coincidence and no more".).
[88] J. Friedman, "Apportionment: Shining the Light of the Day on Patent Damages" (2012) 63 *Case Western Reserve Law Review* 175.
[89] *Metallic Rubber Tire Co v Hartford Rubber Work* 275 F. 315 (2d Cir. 1921).
[90] *P.P. Mast and Co v Superior Drill Co* 154 F. 45, 53 (6th Cir. 1907).
[91] *Celanese International* [1999] R.P.C. 203 at [81].

profit attributable to the infringing matter is the base allocated profits attributable to the invention. He gives a hypothetical example to illustrate this concept.[92] In this example he suggests a situation in which an article with five elements is sold for £1,000, with each of the elements contributing £200. He then suggests that if one of those elements is replaced with another that happens to be an infringing element but the same £1,000 as originally earned is still gained, £200 will still be attributed to that infringing element. This is because the infringing element is still the basis of the profit even though it adds no increased value over the substituted non-infringing element.

In contrast, if the article earns £1,050 because one of the original elements is replaced with an infringing element that adds additional value to the tune of £50, the infringing element will have £250 attributed to it while each of the other four non-infringing elements will have £200 ascribed to them. In both situations £200 appears constant in that in the first situation it is the only sum attributable, while in the second it has £50 added to it. Thus £200 is the base allocated profit, being the minimum sum attributable to the infringing element.

Apparently, this feature of the apportionment approach bears a static, but not dynamic, view of inventions. This is because it is concerned about the gains derived from a patented invention on account of only the patentee and the infringer (i.e. static consideration), but not having regard to other technical contributions solving the same problems (i.e. dynamic consideration), at least before the patent was granted. This likely feature of the apportionment approach appears to be out of tune with the nature and spirit of advancement in technical knowledge. This is because knowledge is cumulative and as such a true measure of profits gained from an infringing application of it can best be determined against the measure of gains that could have been earned before the patent was granted.[93] Thus it is submitted that the "base allocated profits" feature of the apportionment approach is at odds with the cumulative nature of knowledge. An excerpt from Polanyi's writing illustrates this:

> "Ideas usually develop gradually by shades of emphasis, and even when, from time to time, sparks of discovery flare up and suddenly reveal a new understanding, it usually appears on closer scrutiny that the new idea had been at least partly foreshadowed in previous speculations … Invention, and particularly modern invention which relies more and more on a systematic process of trial and error, is a drama enacted on a crowded stage. *It may be possible to analyze its various scenes and acts, and to ascribe different degrees of merit to the participants*; but it is not possible, in general, to attribute to any of them one decisive self-contained mental operation which can be formulated in a definite claim."[94]

Therefore, the flaw of this feature is that a false attribution of profits to the patented invention can arise. This is simply because the patentee is always likely to receive something, however small, so long as technical effect is considered attributable to the incorporation of a patented element even in circumstances where a freely

[92] *Celanese International* [1999] R.P.C. 203 at [58].
[93] See *Cincinnati Car Co v New York Rapid Transit Corp* 66 F. 2d 592 (1933).
[94] M. Polanyi, "Patent Reform" (1944) 9 *Review of Economic Studies* 61, 70–71 (emphasis added).

available non-infringing matter, existing before the patented invention was conceived, would have enabled equal effects. However, if the differential approach is applied this is unlikely to happen. This can also be illustrated using an extension supplied by Laddie J to his hypothetical example stated above. His Honour hypothesised that on the occasion where the infringing article earns £1,050 and a differential approach were applied, the patentee would only be entitled to £50.[95] This is because that sum marks the difference in profits between what could have been earned without infringing (£1,000) and what was earned while infringing (£1,050).

On the occasion where the infringing article earns £1,000, the differential approach will ascribe no profits to the infringement, because it adds no additional value over what could have been earned without infringement. Thus the differential or incremental approach is merely concerned with a comparative assessment of profits that would have been earned with and without infringement, and upon making this assessment the difference in profits so determined is considered the subject for disgorgement. In other words, it is concerned with a dynamic evaluation of the patented invention in relation to other alternative technical solutions available before the patent was granted. This way the pitfalls of the base allocated profits can be avoided.

Uncertainty in parameters for apportionment

Apportionment is always problematic in terms of delineating the value attributable to an infringing element where there are complementary elements comprised in the infringing article that are also responsible for the profits earned.[96] The most common route to apportioning profits is to try extrapolating the value that each element, particularly the infringed element, contributes to the infringing matter. Millett J in *Potton v Yorkclose Ltd*,[97] a UK copyright case, strongly disagreed with this route. He stated that

> "profits ought not in my view be apportioned by reference to evidence of or speculation about the motives of real or hypothetical purchasers or the relative attractions to such purchasers of different aspects of the work".[98]

Dyck has also expressed a similar view. In disagreeing with the use of apportionment, she argues that as certain products are likely to "contain hundreds or thousands of features, it is hard to establish the precise effect of individual component features on consumer demand without creating bias".[99] In *Celanese International*, Laddie J, in furtherance of Millett J's position as expressed in *Potton v Yorkclose*,[100] criticised apportionment based on consumer motivation as follows:

> "Adjusting the apportionment up or down in response to imprecise feelings that one part of a product or process is more or less important or valuable

[95] *Celanese International* [1999] R.P.C. 203 at [58].
[96] See C. L. Kirby, "To Account or Not To Account? An Account" (1991) 13 E.I.P.R. 367; see also L. Bently, "Account of Profits for Infringement of Copyright: Potton Ltd v Yorkclose Ltd and Others" (1990) 12 E.I.P.R. 106.
[97] *Potton v Yorkclose* [1990] F.S.R. 11 Ch D.
[98] *Potton v Yorkclose* [1990] F.S.R. 11 at 19.
[99] P. Dyck, "Beyond Confusion—Survey Evidence of Consumer Demand and the Entire Market Value Rule" (2012) 4 *Hastings Science and Techology Law Journal* 223.
[100] *Potton v Yorkclose* [1990] F.S.R 11.

> than another will add another layer of unpredictability to an exercise which is already difficult enough. It is also likely to result in the account being burdened with evidence directed to flattering or denigrating the relative merits of different parts when, as a matter of commercial reality, the customer does not really distinguish one part from another."[101]

Upon doing so, he moved on to querying apportionment undertaken on the basis of the relative value of components. He considered this approach to be unsustainable and indefensible because it treats each component as though it bore an existence or independent value of its own. His Honour said:

> "Where the part can not be severed from the rest and sold or exploited on its own it acquires its value by reason of its co-operation and interaction with the other parts."[102]

Millet J in *Potton Ltd v Yorkclose Ltd*[103] and Laddie J in *Celanese International*[104] identified an alternative route to determining an apportionment, known as "cost-based" apportionment. This approach apportions profits in proportion to the costs or expenses attributable to the infringing element in entire infringing enterprise. Laddie J identified a major shortcoming of cost-based apportionment, in that "a particular part of a process or product may be very cheap but add a lot of value or vice versa".[105] He did, however, go on to defend the approach, saying:

> "The distribution by costs approach does not set out to distinguish between relatively critical and relatively trivial parts of the whole but it does provide a measure of the base allocated profits attributable to the part in issue."[106]

As already stated, his Honour went on to apply cost-based apportionment in *Celanese*. First, his Honour found that 0.6 per cent and 0.3 per cent, respectively, of the capital expenditure of the two infringing plants (Plants A4 and A5) were expended on the infringing matter (i.e. the patented guard bed). On the basis of this finding, his Honour proceeded to apportion profits. Thus 0.6 per cent and 0.3 per cent of the profits earned respectively from Plants A4 and A5, during the period of infringement, were awarded to *Celanese International Corp*, the patentee. However, the circumstances of the occasion will determine whether the cost-based approach will be used in preference to the value-based route, for, as Millett J said, it should be used "in absence of some special reason to the contrary".[107] In a recent New Zealand copyright case, *ABB v NZ Insulators*,[108] the court considered both routes to apportionment to be valuable, not mutually exclusive, and neither superior to the other. The court, however, adopted the value-based route because it considered it most convenient and apt on the occasion.

[101] *Celanese International* [1999] R.P.C. 203 at [57].
[102] *Celanese International* [1999] R.P.C. 203 at [57].
[103] *Potton v Yorkclose* [1990] F.S.R 11 at 19: "A better guide is likely to be provided by ordinary accounting principles whereby, in absence of some special reason to the contrary, the profits of a single project are attributed to different parts or aspects of the project in the same proportions as the costs and expenses are attributed to them."
[104] *Celanese International* [1999] R.P.C. 203.
[105] *Celanese International* [1999] R.P.C. 203 at [55].
[106] *Celanese International* [1999] R.P.C. 203 at [81].
[107] *Potton v Yorkclose* [1990] F.S.R 11 [1990] F.S.R. 11 at 19.
[108] *ABB v NZ Insulators* (2007) 74 I.P.R. 172.

The uncertainty in the basis of apportionment (i.e. the competition between costs or components as parameters) further mars the doctrinal determinacy of the approach. But, even if the basis of apportionment were certain, an accurate attempt at apportionment is inherently impossible. As Justice Learned Hand correctly observed in *Cincinnati Car Co v New York Rapid Transit Corp*[109]: "It is generally impossible to allocate quantitatively the shares of the old and the new."[110] An unfortunate consequence of apportionment, given its likely uncertainties particularly with regard to burden of proof, is that it could create opportunistic advantages or a windfall for either the patentee or the infringer. This is because, as aptly reasoned by Justice Learned Hand:

> "The party on whom that duty falls, will usually lose. If the patentee is required to assess the contribution of his invention to the profits, he will find it impossible; vice versa, if this is demanded of the infringer."[111]

The likely features of the differential approach

The differential approach also has two main features, these being information costs and unstable outcomes. These features will now be turned to in detail, individually.

Information costs

Information costs in this context may be described as the costs of obtaining, understanding and scrutinising information. As highlighted in *Elwood v Christie*,[112] the natural implication of this approach is a web of information costs such as the "burden of proving their availability, their utility and workability in conditions similar to those under which the work was performed".[113] Allied to this is the question of the period against which that comparison is to be made: is it made pre-infringement or post-infringement? In other words: are we to compare the infringed patents with alternatives available to the infringer pre-infringement or post-infringement?

This likelihood of information costs resulting from the differential approach results from the uncertain and open-ended nature of the standard of comparison adopted in the application of the approach as presently conceived in Canada. In *ADIR v Apotex Inc*,[114] Justice Gagne of the Canadian Federal Court reasoned that the standard of comparison is always between the infringed patent and NIAs that the infringer would most likely have adopted instead.[115] However, in this case the court refused the infringer's suggested NIA because the court reasoned that it was unlikely that the infringer could have adopted any of those alternatives. However, it is important to note that the court, in furthering the reasoning distilled in the growing Canadian case law on the approach, did not state any time-limit on comparison. What can be gleaned, however, is that the Canadian courts are more

[109] *Cincinnati Car v New York Rapid Transit* 66 F. 2d 592 (2d Cir. 1933).
[110] *Cincinnati Car v New York Rapid Transit* 66 F. 2d 592 (2d Cir. 1933).
[111] *Cincinnati Car v New York Rapid Transit* 66 F. 2d 592 (2d Cir. 1933).
[112] *Elwood v Christie* [1865] 18 C.B.N.S. 494.
[113] *Reading & Bates Construction v Baker Energy Resources* [1995] 1 F.C. 483.
[114] *ADIR v Apotex Inc* [2015] F.C. 721.
[115] *ADIR v Apotex Inc* [2015] F.C. 721 at [134]: "First I agree with Strayer J in Reading & Bates … that 'the measurement of profits should be between the infringing method actually used and any other method which would most probably have been used'."

concerned about the infringer's operating realities, which would influence both the choice of NIA and time in which this choice will likely be made.

Siebrasse, an avid proponent of the differential approach, appears to defend this standard of comparison. He simply argues that when comparing, we should look at non-infringing alternatives the infringer would have adopted in consideration of his or her "resources, culture, capacity, and strategic plans".[116] His defence for this position is that it is only consistent with causation when the likely option that the infringer would have adopted is an NIA.[117] He posits that an NIA should be selected on the basis of alternatives available and known to the infringer before he/she has made the decision to infringe.[118]

He argues that this view is consistent with the holding of the US Supreme Court in the famous patent case of the *Cawood Patent* (*Cawood*).[119] In this case, the infringer adopted the infringing method of repairing crushed rail ends that saved it huge costs over the manual method it would have adopted. However, unbeknown to the infringer, its best approach was to have simply avoided repairing, as this would have saved it any kind of expense on repairs. In the course of the disgorgement exercise, the infringer advanced, as an NIA with the infringed patent, the fact that it could have avoided repairs altogether. This is in view of the fact that the more efficient the NIA, the lesser the disgorgement to be made. It follows therefore that where the NIA is more efficient than the infringed patent, the likelihood of disgorgement of any sum at all might be extinguished.[120]

The US Supreme Court in *Cawood* limited the NIA to the manual process that the infringer had known prior to infringement. This was on the basis of the consideration that, had the infringing method not been used, the operating reality of the infringer was to have maintained the more expensive manual process. Therefore, to the extent that the infringed method enabled the infringer to avoid applying the expensive manual process, the infringer gained.

Allowing a comparison of alternatives open to the infringer to be taken into account, even when that comparison is limited to technological matters within the infringer's knowledge before infringement, amounts to a rewriting of history.[121] It calls for an invitation to inquire into what *could have* happened, thus causing focus to move away from *what actually happened*. As Hylands has opined, engaging in a comparison of this kind might enable the infringer to "avoid making amends by showing that he might have done otherwise".[122] This is because the more technically efficient the alternative is, the lesser the amount to be disgorged.[123]

[116] N. Siebrasse et al., "Accounting of Profits in Intellectual Property Cases in Canada" (2008) 24 *Canadian Intellectual Property Review* 82.

[117] Siebrasse et al., "Accounting of Profits in Intellectual Property Cases in Canada" (2008) 24 *Canadian Intellectual Property Review* 82.

[118] Siebrasse et al., "Accounting of Profits in Intellectual Property Cases in Canada" (2008) 24 *Canadian Intellectual Property Review* 82.

[119] *Cawood Patent* 94 U.S. 695 (1876).

[120] E. Phelps, "The Time Element for the 'Standard of Comparison' as Used in Accounting for Defendant's Infringing Use of a Patented Invention" (1922) 5 *Journal of the Patent and Trademark Office Society* 219; see also A. Morsell, "The Burden of Proof in Accounting Proceedings in Patent Suits" (1912) 35 *Annual Report of the American Bar Association* 890.

[121] *Celanese International* [1999] R.P.C 203 at 215.

[122] H. Johns, "The Time of Availability of the Standard of Comparison for Use in Accounting for Profits from Patent Infringement" (1921) 4 *Journal of the Patent and Trademark Office Society* 259, 266.

[123] Johns, "The Time of Availability of the Standard of Comparison" (1921) 4 *Journal of the Patent and Trademark Office Society* 259, 266.

A salient point raised by Hyland is that as science and technology advances, subsequent developments are likely to supersede preceding ones in value, but that this does not necessarily factor into the sum of profits derivable from the former or latter.[124] Therefore, that an alternative exists does not necessarily reflect the sum of profits that could have been earned if it were adopted over another or in fact, the infringed patent.

Hyland also addressed the judicial position expressed in *Columbia Wire v Kokoma Steel & Wire Co*[125] on comparison, disagreeing with the court's reasoning. In that case, the court reasoned that that it is only fair to consider alternatives in the manner similar to that suggested by Siebrasse. The *Columbia Wire* court and Siebrasse consider that there might be a chance that an infringer could have honestly believed it was entitled to use a patented matter, and would have avoided infringing if properly informed about the validity or existence of the patent. On the basis of this, the court reasoned that it was fair to consider alternatives that the infringer might have adopted, because ignoring them would lead to holding the infringer accountable to the patentee for more than the patentee is entitled to, thus causing the infringer to be extorted. To this Hyland has countered that the value of a patent does not necessarily represent the measure of profits derivable from it.[126]

Hyland opined that the farthest we should go in comparing is with alternatives available to the infringer before the patent was issued.[127] This is because the infringer is likened to a trustee, and it will be contrary to the rules of equity to allow the infringer to reduce liability by showing that it could have done otherwise when in fact that alternative might not have been pursued.[128]

Hyland's position that the NIA should at least be an alternative available to the infringer before the infringed patent was granted accords with that of Kekewich J in *Siddell v Vickers*, discussed above. It is submitted that it is the optimal standard for comparison in the application of the differential approach. One reason for this is that is facilitates an economy on information costs as it avoids a morass of difficult inquiries. For example, it avoids questions on the alternative technologies available to the infringer in the light of their peculiarities, bordering on their knowledge and resources to exploit such alternatives. It also avoids an inquiry into the substitutive utility and workability of alternatives within the infringer's reach before infringement. Rather, it simply focuses on the differential profits derivable from viable technical alternatives available before the infringed patent was granted, as against those actually derived from infringement.

[124] Johns, "The Time of Availability of the Standard of Comparison" (1921) 4 *Journal of the Patent and Trademark Office Society* 259, 274–275.

[125] *Columbia Wire v Kokoma Steel & Wire Co* 194 F. 108 (1911).

[126] Johns, "The Time of Availability of the Standard of Comparison" (1921) 4 *Journal of the Patent and Trademark Office Society* 259, 269

[127] Johns, "The Time of Availability of the Standard of Comparison" (1921) 4 *Journal of the Patent and Trademark Office Society* 259, 269.

[128] Johns, "The Time of Availability of the Standard of Comparison" (1921) 4 *Journal of the Patent and Trademark Office Society* 259, 269 ("A trustee might appropriate a bond of his cestui and by marginal speculation reap profits many times the value of the property converted, but no court of equity would say the cestui should be reimbursed only according to the value of the property at the date the trust was breached.").

Likelihood of unstable outcomes

The financial outcomes of a comparative approach can be substantially indeterminate. Laddie J identified this and tried depicting it using an extension of the hypothetical set of facts given above, on the concept of base allocated profits.[129] In furtherance of his Honour's hypothetical illustration, he submits that on the occasion where £1,050 is earned from selling the hypothetical article, £50 might be considered the differential gain from applying the infringing element. However, in another scenario where £1,050 is earned, the court may come to the conclusion that the adoption of the infringing element causes efficiencies to be lost or that the other non-infringing elements could still have attained the same outcomes, but the infringing element only provided appreciable added value. For this, the court may not attribute the whole £50 to the infringing element, but a part of it. Finally, the court might consider it right and fair to distribute the differential gain of £50 equally among the five elements, including the infringing one, such that £250 is attributed to each supporting component or element.

On the surface it would appear that this is a feature that is in substance similar to the effects of the "uncertain parameters" of the apportionment approach. However, both are different from each other, and any such likelihood of similarity could only be wrought by circumstance. This is because the usefulness of the apportionment approach is circumscribed to cases where profits are considered to have been earned from a purely accounting standpoint. It is of no use where profits have not been made in that sense. The apportionment approach operates to apportion profits gained (in the accounting sense) among the various factors or elements employed towards an infringing engagement. The differential approach, being more holistic and tied to the economic or holistic definition of profits, is concerned with how an infringing activity has improved the economic conditions or circumstances of the infringer. Thus it operates on the basis of the question: "but for the infringement, what would the economic conditions or circumstances of the infringer have been?". It answers this question by taking account of other technical alternatives that the infringers could have availed themselves of to ascertain what the infringer really gained.

Admittedly, the supposed indeterminacy of the financial outcomes of the differential approach, as highlighted by his Honour, may be taken, at first blush, as a flaw in the approach. However, that should not be the case. This is because different facts would inspire different considerations in the mind of judicial arbiters, and, as such, the measure of gains from an infringing activity would always be perceived differently. The merit and superiority of the differential approach over the apportionment approach lies in its distinct focus on the incremental gains or advantages derived by the infringer over NIAs available before infringement.

The superiority of the incremental/differential approach in the patent law context

Before justifying a case for the differential approach, it is important to provide signposts regarding how it should be applied. These signposts relate to the two

[129] *Celanese International* [1999] R.P.C. 203 at [58].

major likely situations that could occasion an infringer's gains from infringement. One is where the infringed patent accounts for the entire economic gains made by the infringer, and the other is where the infringed patent is partly responsible for the economic gains derived:

1. **Where the infringed patent is responsible for the entire profits:**

   Let it be assumed that an infringer earns $25 per infringing item sold which could not have been possible but for the infringed patent. In other words, NIAs available to the infringer before the grant of the patent would not have enabled the infringer earning such profit. Going by the differential approach, the entire $25 profit per item sold would have to be disgorged to the patentee. However, even if the entire profit is attributable to the infringed patent, but the infringer could still have earned same by applying an NIA, then in such a situation the profit cannot fairly be attributed to the infringed patent. This is because the infringer did not gain any incremental advantage from the patent which could not have been gained by applying an NIA.

2. **Where the infringed patent is partly account for the infringer's gains:**

   if an infringer earns $25 per infringing item sold, but by applying the best NIA possible the infringer could have earned $20, then the incremental profit gained from infringement is $5. This again, is on the assumption that NIAs are those available before the grant of the infringed patent. On such an occasion, then, only $5 per infringing item sold would be disgorged to the patentee. This is so because $5 represents the incremental monetary value gained by the infringer. Again, if, however, the incremental $5 profit could have been secured by applying an NIA, then it will not be fair to attribute it to the infringed patent.

It is submitted that the incremental or differential approach has merits, in the patent context, over the apportionment approach. This is because it is suitable to the utilitarian nature of patent law policy, which is essentially to reward inventors for their incremental contributions to industrially applicable technological and scientific contributions to knowledge. Being centred on the nexus between the incremental values of the infringed patent over NIAs and the economic advantage gained by the infringer, this approach reflects the purpose of patent law policy.

Also, this method avoids the problematic difficulties of apportionment, and the false attribution of profits to the infringed patent due to the base allocation profits concept can be obviated. Also, the pitfalls of the incremental approach on the basis of information costs can be avoided. Thus the information costs that can assail courts when determining the infringer's next best non-infringing alternative, and the problem of unstable outcomes in the computation of profits, will be easily eschewed.

Another advantage of the method is that by avoiding a cleavage between patent law policy and enforcement rules, the incremental/differential approach avoids opportunism.[130] It avoids patentee opportunism, by ensuring that the patentee is able to receive no more than the value of his or her contribution to technical knowledge as represented in the infringer's gains. Against infringer opportunism, it prevents the infringer from reducing the value of the patent invention and thus reducing the sum to be disgorged, by claiming that he or she would have taken advantage of any of the more efficient alternatives available to him or her before infringement. The effect of this standard of comparison is to cap the likely range of NIAs to be taken into account at those available before the patent was granted. Correspondingly, what the infringer is made to give up approximates a true measure of the profits gained from the infringement.

Conclusion

The foregoing discussions in this article have sought to address the vexed question: "what standard or criterion is best adopted in quantifying the measure of profits or economic advantage(s) derived by an infringing activities in the application of the disgorgement remedy?". Two major issues have been central to the discussions: the controversy over the definition of profits and the competition between the apportionment and differential or incremental approaches as the most suitable method for quantification. These two issues, as shown in the analysis of this article, are connected. An accounting-based definition of profits is linked with an apportionment, while an economic or holistic conception of profits is often linked with the differential or incremental approach. Having shown the superiority of the economic or holistic conception of profits over the accounting-based one, and the superiority of the differential or incremental approach over apportionment, it is submitted that the most suitable criterion for ascertaining the quantum of disgorgement is by:

1. treating as profits both monetary and non-monetary economic advantages derived from the infringement; and
2. ascertaining the gains attributable to the infringement on the basis of a comparison between the patented invention and what existed in the prior art before the patent was issued.

Thus, as regards definition of profits, to allow the infringer to account for only the monetary gains derived from an infringing activity, such that if no profit were made nothing would be disgorged to the patentee, would fly in the face of economic reasoning and the purposes of the patent system. As such it would only be in tune with the realities of the patent marketplace that the monetary equivalence of whatever is derived without the patentee's consent from an infringing activity should be returned to the patentee. Also, concerning causation or quantum of disgorgement, it is only fair and in tune with the nature of the patent system that the incremental or differential profits derived from infringement should serve as a proxy or representation of the invention's marginal or incremental value over technical alternatives available before it was granted. This is in consideration of

[130] H. Smith, "Why Fiduciary Law Is Equitable" in A.S. Gold and P.B. Miller (eds), *Philosophical Foundations of Fiduciary Law* (Oxford: Oxford University Press, 2014), pp.261–280. (Smith defined opportunism "as consisting of as behavior that is technically legal but is done with a view to securing unintended benefits from the system and these benefits are usually smaller than the costs they impose on others".)

the fact that the patent system is aimed at rewarding inventors for the incremental value of their technological and scientific contributions.

Book Reviews

Law of Insolvent Partnerships and Limited Liability Partnerships, by Elspeth Berry and Rebecca Parry, (London: Wildy, Simmonds and Hill Publishing, 2015), lxxiv + 670pp., hardback, £145.00, ISBN: 9780854901609.

Partnership insolvency (in the broadest sense of the concept) is a complex subject, and the authors of this impressive commentary are to be admired for their attempts to illuminate it. They bring to bear their considerable expertise in the twin component areas of partnership law and insolvency law. It would not be possible to provide a credible analysis of this most difficult of subject areas without that diversity of intellectual skills. That combination of talents works most effectively in this particular publishing venture. The text runs to some 670 pages, which itself is an indication of the technicality of the subject area.

The commentary is well written, in a style that will facilitate its use by practitioners, but in a manner that scholars will find satisfying. It covers the law as it stood in April 2015. But this edition does review the potential impact of the Deregulation Act 2015 and the Small Business, Enterprise and Employment Act 2015 insofar as the relevant information about the application of these new statutes to insolvent partnerships was available. The mass of legislative material governing the various regimes applicable to "partnership insolvency" (in the widest sense of the term) is made more palatable by illuminating reference to case law and appropriate legal literature. This text will be the primary source of reference for those keen to study this most difficult of insolvency-related subjects, though a second edition will be required shortly once the Insolvency Rules 1986 are replaced by the putative Insolvency Rules 2016. That said, any changes are more likely to be of form rather than of real substance. There is also the question of addressing the new debtor-induced bankruptcy procedure introduced belatedly via the Enterprise and Regulatory Reform Act 2013 in April 2016. The changes here relate more to the initiation of bankruptcy procedure, rather than to the substance of its operation.

The various chapters provide a good coverage of the key insolvency regimes that might be applicable to an insolvent partnership or LLP. All of the mainstream corporate insolvency procedures, which are adapted to some degree or another to cope with the various partnership structures, are examined in considerable depth and with expert insight. The question of joint bankruptcy is dealt with in Ch.8. Moving away from a discussion of insolvency regimes, Ch.9 considers relevant aspects of insolvency litigation, whereas the final substantive chapter (Ch.10) considers how the Company Directors Disqualification Act 1986 might come into play.

The main text is well supported by valuable appendices featuring both primary and secondary legislation. Standards of presentation are high. The main text is structured using a numbered paragraph format; this facilitates cross-referencing.

I was greatly impressed by this new text. It should feature in any practitioner's library. Any educational institution having a research mission should also make available a copy to its scholars.

Professor David Milman
Lancaster University

European Insolvency Law: Reform and Harmonization, by Gerard McCormack, Andrew Keay and Sarah Brown, (Cheltenham: Edward Elgar Publishing Ltd, 2017), xv + 487pp., hardback, £95, ISBN: 978-1786433305.

Should we harmonise European substantive insolvency law? If so, to what extent?These are two of the most topical research questions that insolvency scholars, practitioners and law-makers have been addressing in recent times. Convergence of insolvency law is a top priority at the European level. Nevertheless, while the demand for informative and practical work, aimed at both specialist and non-specialist readers, is on the rise, up to now much of the debate has been theoretical.

This book represents an appropriate answer to this plea, to the extent that it might be considered as a modern bible for European cross-border insolvency practitioners. It compares existing solutions with international benchmarks and national and cross-border best practices. In addition, it discusses the possibility of harmonisation and reform measures within the EU.

This cross-cutting analysis of the data gathered from a study of insolvency laws, and of significant decisions from the CJEU and other national courts, addresses the lack of empirical, objective evidence on the variety of substantive statutes within the European Member States. It provides novel material to promote the objectives set out in several European documents, including the Capital Markets Union project, and the European Commission's Single Market strategy. However, it is not simply a "picture of the status quo", but a list of suggestions for achievable changes in European insolvency policy for the near future.

This book comes at a time of increased discussion and critical analysis of insolvency law due to international developments such as Brexit. Its publication is certainly timely, since it represents an informative addition to a primarily theoretical discussion. In contrast to Baird's claim that debates about the law of corporate reorganisations are not debates about facts,[1] this comparative book drives the readers to enhance their knowledge, and considers the opportunity for converging measures at EU level. Additionally, the commissioner of this research, the European Commission, will greatly benefit from the disclosed data in making better-informed decisions on proposals for future advancements in this area of law.

The study on substantive insolvency law involved 28 European Member States and two "comparator" countries (the US and Norway). It was commissioned in

[1] D.G. Baird, "Bankruptcy Uncontested Axioms" (1998) 108 *Yale Law Journal* 573.

2015 by the European Commission — DG Justice to a team of scholars and practitioners co-ordinated by the School of Law of the University of Leeds.

The involvement of the European Commission on one side, and the establishment of a team of experienced and renowned insolvency professors and practitioners on the other, simultaneously represents the primary strengths and limits of this project. The devised methodology and structure, and the influential support of the European institutions, has resulted in a rigorous and detailed collection and analysis of the gathered evidence. At the same time, its European focus has resulted in the adoption of otherwise questionable key assumptions, such as that harmonisation of insolvency practices is beneficial for the economy and/or should be pursued. It has also determined a limited analysis of alternative strategies to harmonisation, such as the simplification of administrative burdens in insolvency procedures, or the promotion of informal and contractual solutions to corporate or personal distress.

This study covers some of the most topical subjects where substantive harmonisation of insolvency law has been envisaged and recommended, at both European and international levels. They range from corporates to individual practices, from the role and treatment of the main stakeholders in corporate reorganisations (i.e. directors, insolvency practitioners and creditors) to granting a second chance to unfortunate entrepreneurs, and the debt discharge to over-indebted consumers.

However, its authors went well beyond the data-collection approach adopted in several similar reports and papers. In carrying out this comparative analysis, they also offer a "spaccato" of some of the most controversial issues in current insolvency practice. This is the case, for instance, of the analysis of how the notion of "insolvency" differs from country to country, and how this lack of uniformity affects the duties of directors, as well as the feasibility for a company to reorganise its business and operations at an earlier, rather than later, stage, and/or in cross-border cases.

The same considerations apply to numerous other pivotal concepts, such as the scope, relevance and recognition of the "absolute priority rule". Finally, the book also delves into the strengths and limits of other empirical databases and studies, including the World Bank *Doing Business* project. Consequently, this study adopts a methodology that addresses the primary criticism of the *Doing Business* approach, namely its focus on the law in books rather than the law in action. It is not therefore surprising to observe that—with reference to some jurisdictions and specific issues—the book is silent on evaluating existing policies or providing reform recommendations owing to the lack of available evidence/sources.

The book is divided into eight substantive chapters. Seven of them deal with issues for which a process of substantive harmonisation of insolvency law has either been envisaged or recommended by international bodies (UNCITRAL, World Bank, European Bank for Reconstruction and Development, and INSOL). The eighth one (which, in reality corresponds to Ch.6 of the book) considers the issue of Member State compliance with the 2014 European Commission Recommendation on a new approach to business failure and insolvency.

It is this chapter in particular which is emblematic of the approach of the authors of this study to the contentious issue of convergence of insolvency law. By building upon considerations recently expressed in the literature as well as in other reports (including the Law Review Committee in Singapore), the authors express their cautious approach towards harmonisation. They recognise and explain why detailed solutions may be more appropriate—at least for the time being—to preserve the differences in history, culture and national economies, as well as in the state of economic development between the Member States (p.229). They also warn about the risk of cosmetic, rather than substantive, statutory harmonisation. For instance, they advise that the role of the judiciary cannot be easily reformed with a stroke of the pen, i.e. by means of a top-down reform of established practices.

The relevance of this book lies in its ode to the concept of balance. While using an illusorily descriptive and explanatory style, this book manages to reach the widest audience of insolvency aficionados, to promote the benefits and highlight the risks of a new, harmonised approach to business and individual failure and insolvency.

In a remarkable scholarly exercise, the authors have managed to co-ordinate a research project which conveys new, factual evidence on pivotal and controversial aspects of insolvency practice. However, they also have offered an honest yet challenging analysis of the evidentiary conclusions. While these policy reflections might have deserved autonomous consideration in a separate chapter, they are aptly expressed throughout the entire work.

To conclude, a wide variety of people and organisations, professionals and charities should look at the results of this study with the careful eye of the interested observer.

Eugenio Vaccari
City, University of London