Edinburgh Student Law Review

2018
Volume III | Issue 3

Sponsors

TURCAN CONNELL

DLA PIPER

Edinburgh Student Law Review 2018

The board wish to give their sincere thanks to the anonymous peer reviewers who have contributed greatly to the production of this issue but who cannot, for obvious reasons, be named.

Foreword to the First Volume by Lord Hope of Craighead

I offer my warmest congratulations to those whose idea it was to institute the *Edinburgh Student Law Review* and to everyone who has been responsible for bringing this issue forward to publication. It is the first student-produced law review in Scotland and only the third in the United Kingdom. It is fitting that the Law School at Edinburgh – a city that was in the forefront of publishing political reviews in the age of enlightenment – should lead the way here north of the border.

Ground-breaking though its publication may be in this jurisdiction, the *Review* follows a tradition that has long been established among the leading law schools in the United States. The editorship of student law reviews in that country is much sought after, as is the privilege of having a paper accepted by them for publication. The stronger the competition for these positions, the higher the standard that is exhibited by those who occupy them. It is well known that their editors are singled out by the Justices of the US Supreme Court and the Federal Appeals Courts when they are recruiting their law clerks. The reflected glory that this produces enhances in its turn the reputation of the reviews. A reference to editorship of this *Review* will not escape notice if it appears on the CV of someone who is applying to be a judicial assistant to the UK Supreme Court. But participation in its publication will be of benefit in so many other ways too.

This is pre-eminently a publication by and for students. Its aim is to enhance standards of thinking and writing about law and to promote discussion among all those who are studying law, at whatever level this may be. Law is pre-eminent among the professional disciplines in its use of words to convey ideas. Thinking and writing about law is an essential part of legal training. So too is the communication of ideas about law, as each generation has its part to play in the way our law should develop for the future. I wish all success to those who will contribute to this project, whether as writers or as editors, and I look forward to the benefits that will flow from making their contributions available through this publication to the wider legal community.

David Hope
March 2009

Editorial

As the Edinburgh Student Law Review turns ten years old, we would like to extend our warmest congratulations to the students who started it, in 2008, and the Law School, which embraced it. It was, at the time, a ground-breaking idea in Scottish legal scholarship, but one rooted in the deep tradition of student law reviews in the United States. At this pivotal anniversary, we would also like to thank all the students and members of staff at the University of Edinburgh who have, throughout this past decade, helped to make the Review a success. The support and continuity provided by our Honorary President, Lord Hope of Craighead, and Honorary Secretary, Dr Andrew Steven, have been invaluable to its progress.

Since its inception, the Review has aimed to provide students at all stages of their academic career with a forum to discuss their views on a wide variety of legal issues. This year's edition is no exception. Having received a record number of submissions, we are able to produce and publish an issue that includes articles in Scots private law, international law, family law, and European Union law. We are also pleased to publish articles written by accomplished criminology students.

We are also very proud that the Review remains a publication entirely run by students, ranging from undergraduates to doctoral candidates. In the past ten years, students at all levels of their studies have demonstrated a willingness to engage with the law, express their views, and contribute to important debates. Equally, they have volunteered their time and energy to make the Review a success. We would like to take this opportunity to thank those who came before us, but also the extraordinary team we have had the privilege of leading this past academic year. Each issue of the Review is a collaborative effort between many individuals, each bringing a vital part to the final publication.

We would like to thank the publications team, which ensured any events were successful, and the finance team, which maintained our relationship with our sponsors, and ensured the Review is financially healthy. We would also like to thank those members of the PhD community who worked as content editors on the current issue and peer reviewed all submissions. And last - but certainly not least - we extend our thanks to the copy editors, who worked very hard, and on a tight schedule, to ensure all articles are of the highest standard. A special thanks needs to be extended to Dr. Tobias Lock, who provided our guest article this year; his take on the topical issue of Brexit should make for an interesting read.

As this is our first year as co-editors-in-chief, we have relied on the knowledge and experience passed on by our immediate predecessors: Alisdair MacPherson and Alasdair Peterson. We have endeavoured to maintain their high standards.

Finally, we would like to thank our sponsors, whose support has allowed the Review to go from strength to strength.

It has been a great ten years, here's to ten more!

Andrew Sweeney & Dionysios Pelekis
Co-Editors-in-Chief
2017-2018

Table of Contents

LOST IN TRANSITION?

Tobias Lock[*]

With less than a year to go until the two-year negotiating period comes to an end, readers of the Edinburgh Student Law Review will no doubt have realised that Brexit is a complex process. A transition period – agreed in principle by the EU and the UK in the draft withdrawal agreement[1] - is considered necessary to ensure an orderly Brexit. The rationale behind it is twofold: nobody expects an agreement on the future relationship between the UK and the EU to be negotiated and ratified by 29 March 2019; and even if it were, there would not be enough time for governments, traders, and individuals to prepare for its entry into force. Given there is no political appetite for an extension of the period for negotiating the withdrawal agreement,[2] a transition based on the status quo of the UK's EU membership is the only way to avoid the EU-UK relationship finding itself in a disruptive limbo with trade on the basis WTO rules and other forms of cooperation either not happening or based on largely out-dated international treaties.

As this contribution aims to show, a status quo transition sounds simple but, like all things Brexit, putting it in place will be legally challenging.

One can distinguish four dimensions to transition: the UK-EU dimension; the EU-internal dimension; the UK-internal dimension; and the external dimension.

A. THE UK-EU DIMENSION

It is evident from its inclusion in the withdrawal agreement that the Commission intends the transitional arrangement to be agreed on the basis of Article 50 TEU. It is, however, by no means certain whether Article 50 TEU is the appropriate legal basis for this. It is certainly preferred by the Commission as in procedural terms it is relatively light-touch: it only requires a qualified majority in the Council (and not unanimity, let alone all the EU-27 as parties to the agreement) and the consent of the European Parliament. The choice of the appropriate legal basis is a matter of constitutional significance[3] given that the EU only has limited powers to conclude international agreements on its own; moreover, different legal bases may require different procedures to be followed.

In the Court of Justice's own words 'the choice of the legal basis for a measure must be based on objective factors which are amenable to judicial review'.[4] Article 50 TEU does not mention transition, but only allows the EU to 'conclude an agreement with [the UK] setting out the arrangements for its withdrawal.' Can one consider transition to be covered by the arrangements for the UK's withdrawal on this basis? The argument in favour is that the essential object of transition is to facilitate the UK's withdrawal. Counter-arguments might point to the more specific EU competences on trade and indeed the express possibility to extend the transition period foreseen in Article 50 (3) TEU, use of which would avoid all the problems discussed here.

---

[*] Senior Lecturer, University of Edinburgh Law School

[1] Draft Agreement on the withdrawal of the United Kingdom of Great Britain and Northern Ireland from the European Union and the European Atomic Energy Community, available at: https://ec.europa.eu/commission/sites/beta-political/files/draft_agreement_coloured.pdf [accessed 9 May 2018]; note that all of part four of the agreement dealing with transition is coloured in green, which means that the text has been agreed by both sides.

[2] This would be possible according to Article 50 (3) TEU provided the UK and the Council of the EU (acting unanimously) agreed.

[3] Opinion 2/00 ECLI:EU:C:2001:664, para 5.

[4] Case C-84/94 United Kingdom v Council ECLI:EU:C:1996:431, para 25.

Assuming then that in principle transition can be achieved on the basis of Article 50 TEU, does transition have to be time-limited? It would seem clear from the wording of Article 50 that it is confined to agreeing the terms of withdrawal and cannot serve as a basis for a (permanent) future relationship. This would suggest that a time-limit must be included and indeed, draft agreement provides that the transition period end on 31 December 2020.

This results in the practical problem that the EU and the UK will again end up negotiating under time-pressure. The draft withdrawal agreement does not feature an option for extension. It is highly doubtful, however, that an extension can be agreed on the basis of Article 50 TEU after the withdrawal agreement has entered into force. After all Article 50 TEU provides the competence basis to conclude an agreement on the withdrawal of a Member State; but once the transition period has started the UK will no longer be a Member State. An extension would have to be done on a different basis and probably as a mixed agreement. The express possibility of a time-limited extension would therefore be preferable although even in that case it would be unclear whether it could be agreed on the basis of Article 50 TEU in advance as e.g. an unlimited number of extensions might be considered too 'permanent'.

In terms of content, Article 122 (1) of the draft agreement adopts a catch-all approach by stipulating that 'unless otherwise provided Union law shall be applicable to and in the [UK] during the transition period.' EU law will produce the same legal effects as before, i.e. it will have primacy and some EU law will be directly effective. At the same time, the UK will no longer be a Member State and will thus no longer participate in the EU institutions.[5] The Union's institutions – in particular the Commission and the Court of Justice – will continue in their roles as supervisors and enforcers of Union law in the UK.[6]

B. THE EU-INTERNAL DIMENSION

From an EU-internal perspective transition raises three key questions: the status of the withdrawal agreement under EU law; whether its provisions will have direct effect in the EU; and how it will be interpreted.

Like every other international agreement concluded by the EU, the withdrawal agreement, and thereby the transition provisions, will occupy a status between EU primary and secondary law. This means that the agreement must comply with substantive EU law (e.g. the EU Charter of Fundamental Rights), but also with key constitutional principles, particularly those concerning the integrity of the EU's constitutional framework. Compliance is reviewable by the ECJ and any deviation from the status quo risks being found incompatible with the overall Treaty framework.

It will in practice be important for individuals in the EU-27 to be able to rely directly on the transitional arrangements. There is a long line of ECJ case law[7] confirming that agreements concluded by the EU are capable of having direct effect: (a) if the nature and broad logic of the agreement do not preclude it; and (b) if the concrete provision is sufficiently clear, precise and unconditional.[8] Some EU agreements expressly exclude such effect, for instance Article 30.6 CETA.

The draft agreement is somewhat unclear in this regard. It requires that applicable Union law shall 'produce in respect of and in the UK the same legal effects as those which it produces within the Union and its Member States', but it is silent on the effects of the agreement in EU law. Applying the

[5] Article 123 (3) of the draft.
[6] Article 126 of the draft.
[7] Starting with Joined Cases 21-24/72 *International Fruit Company* ECLI:EU:C:1972:115.
[8] See e.g. Case C-240/09 *Lesoochranárske zoskupenie VLK v Ministerstvo životného prostredia Slovenskej republiky* ECLI:EU:C:2011:125, para 44.

above test set out by the ECJ, there should be little doubt that EU law – whether applied as such or through the prism of the withdrawal agreement – is capable of having such effects. Nonetheless, an express clarification in the withdrawal agreement would be helpful. Otherwise there remains a risk that the ECJ will not consider the withdrawal agreement to be capable of having direct effect, which would lead to an imbalance in the EU-UK relationship.

Finally, it is unclear how cross-cutting principles of EU law, such as 'mutual trust'; the 'duty of loyalty'; or 'effet utile' will apply under the transitional arrangement. On the one hand, the Commission tries to ensure a continuation of the status quo. On the other, these principles stem from the supranational nature of EU law, which the UK has decided to abandon. What *effet utile* would one invoke when interpreting EU law as it applies during the transition period? Surely, the purpose of the transition period is to facilitate the end of UK membership and is therefore not the same as that of the EU Treaties, which aim at closer integration.

C. THE UK-INTERNAL DIMENSION

The draft withdrawal agreement requires the UK to achieve the following in regard of transition: continued direct effect and primacy; a legal basis for the transposition of Directives; continued jurisdiction of the ECJ.

As the UK is a dualist state, this means that it will need to enact legislation to achieve this. At present, this role is performed by the European Communities Act 1972. That Act, however, will be repealed by the European Union (Withdrawal) Act that is currently making its way through Parliament. One option would be to delay the repeal of the 1972 Act and amend it so that it captures the transitional rules as well.

But this would seem politically improbable as it overlooks the totemic nature of the Act for those in favour of Brexit, so its repeal seems inevitable. Rather oddly, the UK Parliament will then have to re-enact almost identical provisions for transition. The Government has already announced a new Withdrawal Agreement and Implementation Bill,[9] which is likely to perform this role.

The relationship between this new Act and the European Union (Withdrawal) Act will need to be clarified. The latter cannot simply be delayed as it will repeal the European Communities Act and therefore the category of 'retained EU law' that it introduces will still be needed to ensure that there are no gaps in the legal system. Nonetheless, the Withdrawal Agreement and Implementation Act will need to make sure that ministers cannot use their powers of amendment[10] under the European Union (Withdrawal) Act since during transition the UK must remain aligned with EU law as if it were a Member State.

Additionally, the devolution acts will need to be amended. The powers of the devolved legislatures and executives are limited in that they currently cannot act in a way that is incompatible with EU law.[11] As EU law is currently defined as 'all those rights, powers, liabilities, obligations and restrictions from time to time created or arising by or under the Treaties' and 'all those remedies and procedures from time to time provided for by or under the Treaties', an amendment will be needed. This amendment would need to be coordinated with the amendment to the same provisions envisaged by the European Union (Withdrawal) Bill.

---

[9] https://www.gov.uk/government/news/new-bill-to-implement-withdrawal-agreement.
[10] These powers are highly controversial, see e.g. House of Lords, Constitution Committee, 9th Report of Session 2017-19, European Union (Withdrawal) Bill, HL Paper 69, available at: https://publications.parliament.uk/pa/ld201719/ldselect/ldconst/69/69.pdf.
[11] See for Scotland ss. 29 and 57 of the Scotland Act 1998.

Furthermore, there is no guarantee that the UK Supreme Court will continue to accept the effects of EU law during the transition period as unconditionally as it has hitherto done.[12] While this presents only a small uncertainty, the Withdrawal Agreement and Implementation Bill should try to be as clear as possible on this point in order to dispel any doubts that Parliament may not have wanted a continuation of the status quo also under domestic law.

D. THE EXTERNAL DIMENSION

The final set of hurdles in the way of an effective transition relate to its external dimension.

The draft agreement has resolved one contentious question – whether the UK will be allowed to negotiate or even conclude free trade agreements with non-EU countries during transition – in favour of the UK. The UK is now expressly entitled to 'negotiate, sign and ratify international agreements entered into in its own capacity in the areas of exclusive competence of the Union, provided those agreements do not enter into force or apply during the transition, unless so authorised by the Union'.[13]

The bigger question in practice, however, is whether the UK will still be able to benefit from current EU free trade agreements after Brexit. The draft agreement expressly stipulates that the UK 'shall be bound' by them.[14] A unilateral commitment to be bound would certainly be welcomed by third countries such as Canada or South Korea, but there would be no incentive for them to reciprocate this without anything in return. An alternative formulation, such as a 'continue to apply' clause, would not change anything, however: after all, the EU and the UK cannot legally agree that a third country should be bound something agreed between the two. This is the rule of *pacta tertiis nec nocent nec prosunt* which is a general principle of international law that can also be found in Article 34 of the Vienna Convention on the Law of Treaties.

There are some rays of hope, however, at least as far as the transition period is concerned. Article 15.15 of the EU-South Korea FTA for instance states that it 'shall apply, on the one hand, to the territories in which the Treaty on European Union and the Treaty on the Functioning of the European Union are applied and under the conditions laid down in those Treaties, and, on the other hand, to the territory of Korea'. And: 'As regards those provisions concerning the tariff treatment of goods, this Agreement shall also apply to those areas of the EU customs territory not covered by [the previous paragraph]'.

One could argue that during transition the EU Treaties will continue to be applied in the UK; and the UK will also remain in EU customs territory during that time. This would mean that EU free trade agreements containing such a clause would remain fully operational during transition. At the same time there is no guarantee that this interpretation will be shared by the third country concerned.

A final aspect of the external dimension concerns the EU's Common Foreign and Security Policy (CFSP). It is not exempt from the general rule that Union law is applicable to the UK contained in Article 122 (1) of the draft withdrawal agreement. This means that the UK will generally be bound by decisions adopted under the CFSP, notably sanctions decisions. The CFSP is usually considered the last vestige of intergovernmentalism in the EU. The fact that the UK promises to be bound by decisions taken under the CFSP without having had any influence on them is, therefore, in itself remarkable.

---

[12] Starting with the famous decision by the House of Lords in *Regina v Secretary of State for Transport, Ex parte Factortame Ltd. and Others (No. 2)* [1991] 1 AC 603 (HL).
[13] Article 124 (4) of the draft agreement.
[14] Article 124 (1) of the draft agreement.

The agreement contains a safety valve for the UK, however. Article 124 (6) of the draft agreement provides for the possibility of the UK making a formal declaration 'indicating that, for vital and stated reasons of national policy, in those exceptional cases it will not apply the [CFSP] decision.'

This means that while the UK will no longer be in a position to block a CFSP decision – as Member States are – it will at least be able to avoid being bound by one that contradicts 'vital and stated reasons of national policy'. In doing so, the UK must 'refrain from any action likely to conflict with or impede Union action based on that decision.[15]' This would seem to be a sensible compromise between the protection of the UK's interests as a non-Member State wishing to pursue an independent foreign policy and the EU's interest of no longer wanting the UK being able to interfere with its foreign policy goals.

At the same time it raises the question of how far this exception is judicially reviewable. There is nothing in the draft withdrawal agreement to say that it is not. This again is noteworthy given that Article 275 TFEU excludes the Court of Justice's jurisdiction over the CFSP. Given that the withdrawal agreement as such is not part of the CFSP even where it deals with its application to the UK, one could argue that its jurisdiction to decide on whether the UK has violated its obligations under the withdrawal agreement – provided for in Article 163 of the draft though not yet agreed – is unaffected.

E. CONCLUSIONS

This contribution has tried to demonstrate the considerable complexities of the attempt to create a period allowing for a relatively smooth transition from EU membership to life outside the EU as a third country. And it is all too easy to get lost in transition. In a way, transition is a miniature version of the complexities of Brexit as a whole. And many of the difficulties of both Brexit and transition could easily be avoided. Transition would not be necessary if the Article 50 period were extended; and many of the hiccups of the Brexit process could be avoided by careful planning and a dose of realism. There will be at least one group benefitting from these added difficulties: current and future lawyers.

[15] Article 124 (6) of the draft agreement.

NEGOTIATING A MODERN RETAIL LEASE: CHALLENGES FACED BY THE 21ST CENTURY LANDLORD

Rebecca Muir[*]

A. INTRODUCTION

One of the most significant challenges faced by retailers operating in Scotland during the twenty-first century has been the development of a strong e-commerce market. Many industry experts have heralded the death of the high street due to online shopping, with its benefits usurping the need for traditional high street stores.[1] For traditional retailers who operate primarily from physical stores, consumer demand for omnichannel retailing has resulted in increased investment in online operations and decreased investment in real estate.[2] Although these changes have allowed retailers to meet the demands of the modern shopper, decreased investment in real estate has led to 9% of retail property in Scotland being vacant.[3] In light of this high level of vacancy, it is suggested that retailers require a

[*]Current DipLP student at the University of Edinburgh. The author would like to thank Lorna Richardson and Paul Muir for their assistance.

[1] D Jinks, "2030: Dead End for the High Street" (2017), available at https://www.parcelhero.com/blog/news-updates/2030-dead-end-for-the-high-street.

[2] Omni channel retailing attempts to provide a seamless shopping experience for consumers across physical and digital selling platforms. TLT Solicitors, "Going for Growth: Retail Growth Strategies in 2017" (2016), available at *http://www.tltsolicitors.com/insights-and-events/hot-topics/retail-report-2017/*.

[3] British Retail Consortium, "Vacancy rate remains stable though footfall decline deepens" (2017), available at *http://brc.org.uk/news/2017/vacancy-rate-remains-stable-though-footfall-decline-deepens*.

flexible property portfolio that enables them to respond to changes in consumer behaviour caused by the continuing growth of e-commerce.[4]

The purpose of this article is to consider how landlords can offer the flexibility that retailers now require from their property portfolios. It is argued that lease durations, frequent break options, receptiveness to assignation, subleasing and innovative lease types are all potential solutions to the need for flexibility. This article will also consider the challenges landlords now face because of increased consumer reliance on technology, and furthermore, how technology may change the way we look at traditional leasing concepts such as 'turnover'.

B. ACHIEVING FLEXIBILITY IN RETAIL LEASES

(1) Shorter leases and break options

Retailers operating both online and offline should be reluctant to being tied to physical retail property for long periods, as technological advancement and resultant changes to consumer behaviour could render their physical retail space redundant. Although the average lease length has decreased from twenty-five years to seven years, the exponential growth of technology means that the selling environment could change considerably over the duration of an average lease.[5] In other jurisdictions, leases of commercial premises have a duration of only two years, ostensibly permitting landlords to offer retailers even shorter leases in order to respond to decreased demand for retail property.[6]

There is also underutilisation of tenant break options in retail leases in Scotland, which demonstrates that the sector has yet to adequately respond to the dynamism of the digital age.[7] Tenant break options would allow retailers to terminate the lease prematurely if the selling environment changed and it became impractical for them to continue operating from a physical store.[8] It is possible that short leases that incorporate break options which operate on a monthly, quarterly, or biannual basis are likely to be welcomed by retailers as they allow them to be more responsive to changes in consumer behaviour. However, such frequent break clauses may not be a realistic option. Break clauses result in unpredictability for landlords and it is perhaps unlikely that landlords would be willing to agree to such terms. Ultimately, the likelihood of a landlord agreeing to frequent break clauses will depend on the negotiating strength of an individual tenant and the market conditions within which the negotiations are conducted.

However, regardless of their potential to encourage occupancy, landlords may be reluctant to negotiate frequent break options as they jeopardise their financial security and stability.[9] Instead of having guaranteed income for twenty-five years, short leases with frequent break clauses will only give a guaranteed income for a short term. However, the financial uncertainty that landlords accept when agreeing to break options means that, theoretically, they will be able to charge more rent to compensate for the risk that they are undertaking.[10] At a simple microeconomic level, a low supply of flexible leases means that pioneering landlords will be able to demand more rent for their flexibility as

---

[4] N Crosby, V Gibson and S Murdoch, "UK commercial property lease structure: landlord and tenant mismatch" (2003) 40 Urban Studies 1487 at 1489.
[5] MSCI & Strutt and Parker, "U.K Lease Events Review" (2016), available at *https://www.msci.com/documents/10199/6bb1feeb-b1ee-41b2-ab63-4a4ff7513160*.
[6] K Gerber, *Commercial leases in Scotland: A practitioner's guide*, 3rd edn (2016) para 10.01.
[7] Only 31% of leases in the retail sector have break options. See MSCI & Strutt and Parker, "U.K. Lease Events Review" (2016).
[8] For information on break options see A McAllister, *Scottish Law of Leases*, 4th edn (2013) 211-212; J Rankine, *A Treatise on the Law of Leases in Scotland*, 3rd edn (1916) 527-532; and R Rennie, *Leases* (2015) 141.
[9] B Cooper, "Analysis: Lease expiries will bring retail opportunities" (2015), available at *https://www.retail-week.com/stores/property/analysis-lease-expiries-will-bring-retail-opportunities/5072426.article.*
[10] P McAllister, "Pricing short lease and break clauses using simulation methodology" (2001) 19(4) J of Property Investment and Finance 361 at 367.

retailers will be willing to pay more for a lease that meets their needs and gives them the flexibility needed to respond to the continuous and sporadic growth of e-commerce.[11]

(2) Alienation of the lease

As well as offering shorter leases with more frequent tenant break options, landlords in the digital age can respond to low levels of demand for retail property by being more flexible in their approach to subleasing and assignation. Retailers are experiencing the lowest levels of footfall since 1998 as a result of e-commerce changing consumer behaviour.[12] It is no longer necessary for a consumer to visit a physical store to make a purchase, and those who do choose to visit physical stores increasingly do so as part of a leisure activity.[13] Many retailers are trying to overcome this low footfall by attempting to share their premises with a leisure operator who can provide consumers with an attractive and enjoyable "leisure" environment.[14] The 'big four', Sainsburys, Tesco, Asda and Morrisons, have begun to share their space with leisure operators to increase footfall.[15]

(a) Subleasing

The integration of leisure facilities can be achieved by subleasing part of the retail premises to a leisure operator. Subleasing is where a tenant lets the subject to a third party and acts as its landlord.[16] In Scotland, a tenant is free to sublease if there is nothing in the lease that precludes it from doing so. However, it is common for retail leases to be drafted in a way that limits a tenant's ability to sublease. Thus, in the digital age landlords may be able to encourage investment in retail leases by limiting the restrictions placed upon a tenant's ability to sublet.

As standard, it is quite common for the lease to preclude "partial subletting".[17] In the digital age, however, such a qualification should be avoided unless the retail space is particularly small and it would be impractical to sublet the property.[18] This is because such a qualification would prevent the retailer from subletting part of the property to a leisure operator. Instead of strict preclusions against 'partial subletting', the landlord should ensure that their consent is sought before a tenant makes the decision to sublease. Such limitation means that landlords retain control over their property while tenants are still able to sublet to leisure operators if appropriate.[19]

(b) Assignation

Although some retail tenants are shrinking their property portfolio by subleasing part of the property to a leisure operator, omnichannel retailing and investment in e-commerce means that others may want to make more substantial changes to their property portfolio by alienating the entire premises. Supposing physical retail property remains important in the digital age, the number of stores needed

---

[11] Crosby et al (n 4) 1489.
[12] C Fenech, "The Deloitte Consumer Review: Reinventing the Role of the High Street" (2013), available at *https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/consumer-business/deloitte-uk-consumer-review-role-of-the-high-street.pdf*.
[13] *Ibid*.
[14] 68% of the United Kingdom's top retailers are attempting to share their premises with a leisure operator in order to drive footfall into their store. See TLT Solicitors (n 2) at 12.
[15] S Creasey, "The big four grocers are experimenting with sublets and concessions but will it pay off?" (2016), available at *http://www.propertyweek.com/news/retail/the-%E2%80%98big-four%E2%80%99-grocers-are-experimenting-with-sublets-and-concessions-%E2%80%93-but-will-it-pay-off?/5079492.article*.
[16] For more information on subleasing see Rankine, *A Treatise* 190-199; Rennie, *Leases* 273-300.
[17] Rennie, *Leases* 277.
[18] V Sandberg, "Subletting in commercial leases" (2014), available at: *http://www.longmores-solicitors.co.uk/site/blog/commercial-property/subletting-in-commercial-leases*.
[19] For the law on "consent not to be unreasonably withheld" see *Burgerking Ltd v Rachel Charitable Trust Ltd* 2006 SLT 224 at 228-229 per Lord Drummond Young; and *International Drilling Fluids Ltd v Louisville Investments (Uxbridge) Ltd* [1986] Ch 513 at 519-521 per Balcombe LJ.

to be successful is considerably smaller as a result of customers increased reliance on websites.[20] If the lease is far from reaching its maturity or a break option, a retailer may want to alienate the property by assigning or subleasing the whole subject to a third party. From the author's perspective, retailers should avoid sub-leasing the whole property. It creates a three-tiered relationship whereby the original tenant remains liable to the landlord but is not in possession of any property.[21] Unless a lease imposes joint and several liability on the tenant and his assignees, assignation is a preferable solution for retailers looking to reduce their property portfolio, as it removes them completely from the relationship.[22]

The ability to assign may encourage retailers to invest in leases, as it gives them the flexibility they require to respond to changes in consumer behaviour that lessen the value or utility of physical retail premises. Landlords should avoid strictly precluding assignation and, as with subleasing, they should also ensure that they do not qualify the tenant's ability to assign to a great extent. Historically, landlords may have tried to protect their investment by consenting to assignation only if the prospective third party had a covenant strength "at least as good as the tenant".[23] This is because, unlike subletting, the assignee becomes responsible for all the original tenant's obligations. However, landlords should not always revert to a traditional covenant strength test, but instead take sensible risks and recognise that consenting to assignation of the lease to an innovative and fresh tenant of "weaker" covenant strength is preferable to having a retailer in occupation whose failure to assign may result in too large a property portfolio.[24]

That said, a tenant of weaker covenant strength, though fresh and innovative, can be just as high risk as a struggling but previously financially sound retail tenant. Although a landlord can protect himself from the risks of a less experienced tenant by requiring third party protection such as guarantees, he may still be unwilling to take the gamble. However, even the most cautious of landlords should appreciate that it is not in their interest to stop a struggling tenant from exiting a lease. If a landlord does choose to take a strict approach to assignation to enable them to retain control over who occupies their premises, they can provide the requisite flexibility by ensuring that there are frequent break clauses in the lease. For many struggling tenants, assignation may be a way of reducing an overtly large property portfolio. If this door is closed to the tenant, another one must be left open by the landlord; otherwise they risk their premises being occupied by an insolvent tenant who has been unable to meet the demands of twenty-first century retailing.

(3) Innovative lease types

As well as considering how existing leases can be made more flexible, it is important to consider how more innovative forms of lease may provide landlords with opportunities in an omnichannel retail market. If landlords are unable to obtain a permanent tenant by negotiating flexible leases, they may be able to find a temporary tenant by exploring innovative retailing strategies such as pop-up shops and pop-up windows.[25]

(a) Pop-up shops

Although the flexibility offered by pop-up shops is relevant to all retailers, it is particularly relevant to e-tailers who are adopting omnichannel approaches to selling and are looking to invest more in real

[20] J Brown, H Atherton and C Burnett, "The Rise and Rise of Multi-Channel Retailing. 'Click, Brick, Flick'" (2012), available at https://www.revocommunity.org/documents/get_lob?id=68&age=&field=file.
[21] Rankine, *A Treatise* 190; Rennie, *Leases* 273.
[22] Rankine, *A Treatise* 180-190.
[23] Gerber, *Commercial leases* para 15.08.
[24] J Bourke, "What went wrong with BHS's mega property portfolio?", available at http://www.standard.co.uk/business/what-went-wrong-with-bhss-mega-property-portfolio-a3235001.html.
[25] Pop-up shops and windows are temporary retail premises that can be used to test the market. They normally only exist for very short periods of time.

estate.[26] E-tailers are retailers who operate solely from the online platform and thus have no experience with the physical property market. Pop-up shops can allow them to experiment and test the market before they decide whether to invest permanently in real estate assets.[27] Although not all e-tailers will decide to invest permanently in physical property, allowing them to test the market using pop-up shops may result in landlords being able to appropriate a more permanent lease arrangement. For example, Skinny Dip, an online technology accessories brand, experimented with a pop-up shop in London and have now entered into a ten-year lease with the landlord.[28]

As well as potentially leading to a permanent lease arrangement, pop-up shops can benefit landlords in the digital age by increasing footfall. Pop-up shops can refresh the mix of tenants in a shopping centre or retail area, as their novelty can encourage consumers to visit.[29] A refreshed retail environment can positively impact rental growth. For example, in Brixton Market, an area of London with a large proportion of pop-up shops, market rent increased by 80% between 2009 and 2015. This exceeds rental growth in comparable areas of London and has been attributed to the success of pop-up shops.[30]

If a landlord is unable to find a permanent tenant, he may choose to have a permanent pop-up location with rotating tenants. It follows that the landlord must ensure that each tenant obtains the correct planning permission.[31] Although e-tailers are relying more on pop-up shops to test the market, it is also becoming more popular for restaurants to rely upon pop-up locations to test menus and popularity. However, this dual demand for pop-up spaces creates planning issues in Scotland. This is because there is currently a distinction in planning law between hot food retailers and all other retailers.[32] If there is a change of use between a hot food retailer and a traditional retailer, planning permission must be sought.[33] On a practical level, this rigid planning law may not have a real-world effect as it unlikely that a landlord would lease to both traditional retailers and hot food retailers as the fixtures that they would require are very different. However, it is possible to hire and purchase mobile cooking equipment and therefore it is possible that a landlord with a very basic property may encounter demand from both hot food retailers and other more traditional retailers. Current Scottish planning law would require different types of tenants to seek planning permission, which is a costly and time-consuming process, negating the benefits of a pop-up shop.[34] This position can be contrasted with England. English planning law was relaxed to support the conception of pop-up shops, making it possible for the use of retail properties to be temporarily changed without having to obtain planning consent.[35] Additionally, a change in English planning law has made it possible for the use of land to be changed to "any purpose" for twenty-eight day periods without planning permission.[36] Scottish planning law is not as flexible and does not adequately facilitate the opening of a pop-up location.

---

[26] R Williams, "Amazon opens first physical retail store in time for Black Friday" (2015), available at *http://www.telegraph.co.uk/technology/amazon/11972237/Amazon-opens-first-physical-retail-store-in-time-for-Black-Friday.html*.

[27] B Burgess, "Pop-up retailing: The design, Implementation and five rear evolution of an experimental learning project' (2012) 34(3) J of Marketing Education 284 at 285.

[28] M Hunt, "From clicks to bricks" (2016), available at *http://www.propertyweek.com/insight/market-features/online-retailers-from-clicks-to-bricks/5081533.article*.

[29] *Ibid*.

[30] C Borland, "Pop-ups and street food helping bricks and mortar retail fight back" (2016), available at *http://www.costar.co.uk/en/assets/news/2016/March/Strutts-Pop-ups--street-helping-bricks-and-mortar-retail-fight-back-/*.

[31] A Townsend-Wheeler, "Top of the pop-ups" (2016) 160(21) Solicitors J 18 at 19.

[32] The Town and Country Planning (Use Classes) (Scotland) Order 1997.

[33] Town and Country Planning (General Permitted Development) (Scotland) Order SI 1992/223, Schedule 1 para 3.

[34] Note the impact that Simplified Development Zones, introduced by the Planning (Scotland) Bill 2017, could have upon the need to seek planning permission in certain areas of Scotland.

[35] Town and Country Planning (General Permitted Development) (Amendment) (England) Order 2013, SI 2013/1101, part 4 (D).

[36] *Ibid*. part 4 (B).

This means that, for now, innovative landlords are restricted to offering pop-up premises that cater to one type of tenant. However, the inflexible planning regime in Scotland should not stop pioneering landlords from experimenting with pop-up shops. The twenty-first century landlord should play to his strengths and lease either exclusively to food retailers or non-food retailers. It is hoped that there will come a time when landlords are able to do both, but until Scotland adopts a planning approach similar to that used in England, success can still be found in single use pop-up shops.

(b) Pop-up windows

For some e-tailers, the main purpose of a pop-up shop is not testing the physical retail market but advertising their brand. It is estimated that it can cost a retailer £1 million to remain the top search hit on Google for one year and, as e-commerce grows, it will become increasingly difficult for an e-tailer to become noticed on the internet.[37] As such, there is benefit to attracting consumers outside the online sphere by having a physical presence on the high street or in a shopping centre. Pop-up shops can facilitate this marketing, but if the e-tailer is not looking to invest permanently in physical property, a pop-up shop may not be the most practical solution as they require hiring staff, retail infrastructure and merchandising. An alternative for e-tailers looking to advertise their online brand is a pop-up window. Pop-up windows involve the renting of window space in vacant properties and may present a modern alternative to billboards. In London and New York, online retailer NET-A-PORTER rented window space during London Fashion Week.[38] Using augmented technology, customers were able to take pictures of products displayed in the window and purchase them on the NET-A-PORTER app.[39] If landlords are experiencing high levels of vacancy as a result of low demand for retail property, or have periods of vacancy between negotiated leases, pop-up windows provide a potential stopgap method of generating income with very little expenditure and overhead. The only requirement would be that the window space is located in an area of high footfall, as the main purpose of a pop-up window is to create brand awareness.

The concept of pop-up windows is still growing in the United States and, as such, it may be some time before e-tailers try to source window space in Scotland. If pop-up windows become popular in Scotland, it is likely that they will attract media attention and, as a result, drive an increase in footfall. Not only is this beneficial for the tenant who is looking to develop their online brand, it is also beneficial to the landlord who will be wanting to find a more permanent tenant. The advertisement of the pop-up window location will indirectly publicise the landlord's vacancy, while the footfall increase may attract a more permanent retail tenant. It is also likely to benefit existing tenants at the location.

(c) Drafting of pop-up leases

As pop-up shops and windows are temporary, their duration is unlikely to exceed one year. This means that the lease does not need to be in writing.[40] A lease of such short duration can be implied from the parties' conduct.[41] Despite this, it is likely the parties will require their agreement to be in writing to ensure their obligations are clear.

But should the lease mechanism be used to encompass pop-up shop and pop-up window agreements? Leases are often lengthy and complicated contractual documents, and since pop-up shops and pop-up windows are short term, temporary solutions, parties may be reluctant to enter into

[37] M Hunt, (n 28).
[38] R Thomson, "Net-a-Porter makes first bricks-and-mortar move with pop-up interactive shop" (2011), available at *https://www.retail-week.com/technology/net-a-porter-makes-first-bricks-and-mortar-move-with-pop-up-interactive-shop/5028963.article*.
[39] *Ibid*.
[40] Requirements of Writing (Scotland) Act 1995 s 1(7).
[41] McAllister, *Scottish Law of Leases*, 28; Rankine, *A Treatise* 116; G Paton and J Cameron, *The Law of Landlord and Tenant in Scotland* (1967) 19-21.

complex lease negotiations. Instead, most will be content with a lease that is succinct and conforms to the essential requirements of a lease.[42] One solution available to parties entering into pop-up arrangements is reliance upon the Royal Institution of Chartered Surveyors' Small Business Lease (Scotland).[43] This eight page lease is succinctly drafted and is designed for retail leases under five years that require no rent review. Although legal advice should still be sought before entering into a lease agreement, this standard lease covers all the necessary elements in a concise, easy to understand contract. Though the lease is designed for standard retail premises, its basic drafting means that it could be adapted to capture both pop-up shop and pop-up window agreements.

(d) Pop-up shops and secured lending

The utility of pop-up shops may also be limited if there is a standard security over the property. If a landlord has a standard security over the property, he is likely to be restricted by the standard conditions set out in Schedule 3 of the Conveyancing and Feudal Reform (Scotland) Act 1970. In particular, Standard Condition 6 requires the debtor of a standard security to seek the written permission of the lender each time they agree to let the property. Without variation of this standard condition, it may be difficult and time consuming for the landlord of a burdened property to operate as the landlord of a pop-up shop. They would regularly have to seek the permission of their creditor before entering into a lease with the tenant, and the hassle of doing this may negate the advantages of a temporary lease.

However, despite the utility of pop-up premises being limited by planning requirements and the terms of secured lending, it would be unfair to downplay the massive advantage that pop-up shops can have, both upon landlord's operations and the wider retail community. From a landlord's perspective, they offer temporary occupancy solutions and can lead to more permanent leasing arrangements. From a wider perspective, they can increase footfall in struggling areas and can offer an avenue into physical retailing. The current administrative pitfalls of operating a pop-up location do not outweigh the potential success they can bring to a landlords portfolio and local areas, and it is hoped that Scottish landlords will quickly awaken to the positive impact that these innovative leases can have.

C. RESPONDING TO INTEGRATION OF TECHNOLOGY IN PHYSICAL RETAIL PREMISES

The previous section considered how flexibility is key to the modern retail tenant and how twenty-first century retail landlords must be receptive to their tenant's needs to overcome potentially high levels of vacancy. However, flexibility is not the only thing that is demanded. As shopping has gradually moved from a physical experience to an online experience, retailers have had to respond by integrating more technology into their physical premises to provide consumers with a seamless shopping experience across all selling avenues. Internet connectivity is now vital to a retailer's operations and this has created new issues for landlords and their tenants. The twenty-first century landlord must deal with issues that did not exist fifty years ago, and the drafting of their leases also has to reflect this. Certain lease provisions that have been relied upon by landlords for decades have become unworkable, and lease provisions dealing with twenty-first century issues have had to be introduced to ensure that the lease is able operate in the omnichannel world.

The following section will consider how technology has had a knock-on effect on retail leases. In particular, it will consider how internet connectivity has created new points of negotiating

[42] The essential elements that must be defined in the lease are the parties, the subject, the rent and the duration. See Rankine, *A Treatise* 114-116; McAllister, *Scottish Law of Leases* 34.
[43] Royal Institution of Chartered Surveyors, "RICS Small Business Retail Lease (Scotland)", available at http://www.rics.org/Global/RICS%20Small%20Business%20Lease%20-%20Scotland.cmallett.pdf.

for the parties of a retail lease and how turnover rent has become impractical in the modern retail world.

(1) Internet connectivity

For the majority of retailers, the growth of e-commerce means that the provision of in-store Wi-Fi is crucial. Currently, 72% of retailers offer Wi-Fi connectivity to their customers and an additional 15% planned on investing in customer Wi-Fi.[44] It is likely that retailers are making this technological investment in order to capitalise upon the increased use of mobile technology with Mark Felix, director of online trade at John Lewis, suggesting that mobile technology bridges the gap between online and offline retailing.[45]

(a) Provision of customer Wi-Fi

When retailers occupy retail property for a considerable time, it is likely that they will source their own internet connection so that they can choose the provider and the rate that they pay. However, in some situations, landlords may provide a tenant with pre-installed internet connectivity. For example, in shopping centres it is common for the landlord to provide Wi-Fi to their tenant's customers in order to encourage them to visit the shopping centre instead of shopping online.[46] Landlords may also provide customer Wi-Fi to tenants of pop-up shops and windows, as these outlets are used by e-tailers to develop their online brand and so it is likely that they will rely heavily on internet connectivity. However, it can take considerable time to have commercial Wi-Fi installed in retail property and this can be an obstacle to e-tailers looking to take temporary occupation of retail premises.[47] Although in longer leases, landlords may use rent-free periods to compensate for the time where a tenant is unable to trade from the premises, a rent-free period is unlikely to be agreed for a pop-up shop as this would detract from its short and flexible nature. As such, landlords who are looking to lease their property as a pop-up shop may encourage a prospective tenant by providing Wi-Fi as part of the lease.

If landlords do provide customer Wi-Fi, they should ensure that the tenant is responsible for the cost of its supply, management and maintenance.[48] In shopping centres, this could be achieved by expanding the service charge provision to include "internet connectivity".[49] Service charges are extremely common in shopping centres so it is unlikely a tenant would dispute such an inclusion, especially given the service is one demanded by their customers.[50] The expenditure related to the provision and maintenance of Wi-Fi in pop-up shops could also be recovered from the tenant using an all-inclusive rent that covers the standard bills of retail tenants. Pop-up shops are likely to be used by e-tailers who have no previous experience of physical retailing. An all-inclusive rent with no hidden charges can assist novice e-tailers in establishing whether a move to physical property is within their budget. Although it means that the landlord has greater responsibility for the property, an all-inclusive

---

[44] TLT Solicitors (n 2) at 10.
[45] J Bingham, "Are mobiles changing how we shop?" (2016), available at *http://www.telegraph.co.uk/news/shopping-and-consumer-news/12172230/Are-mobiles-changing-how-we-shop.html*.
[46] Footfall in shopping centres has fallen by 4%. See Z Wood and K Allen, "Where have all Britain's shoppers gone?" (2016), available at *https://www.theguardian.com/business/2016/apr/23/uk-high-street-spending-shoppers-slowdown*.
[47] Centre for Economics and Business Research, "Britain's pop-up retail economy 2015" (2015), available at *https://ee.co.uk/content/dam/everything-everywhere/documents/Pop-Up%20Economy%202015.pdf*.
[48] P Troughton, "The evolution of the retail landlord" (2015), available at *http://www.propertyweek.com/opinion/the-evolution-of-the-retail-landlord/5077121.article*.
[49] For information on service charges see Rennie, *Leases* 541-551; McAllister, *Scottish Law of Leases* 324.
[50] Provision of customer Wi-Fi can be of benefit to retailers operating in shopping centres, as 35% of consumers want free Wi-Fi connections as standard when visiting retail premises. See YouGov, "Innovations in Retailing 2015 Report" (2015), available at *https://yougov.co.uk/news/2015/05/22/consumers-want-free-wi-fi-store-improve-shopping-e/*.

rent adds to the attractiveness and simplicity of such a short lease and may encourage e-tailers to invest long term in physical property. However, unless the lease is extremely short, all-inclusive rents should be avoided in favour of services charges. This is because all-inclusive rents require the landlord to accurately predict the service costs in advance.[51] If the lease is longer than a year, it is possible that the cost of providing and maintaining the internet connection could increase and the landlord left financially liable.

In addition to the landlord ensuring that the cost of Wi-Fi is covered by a service charge provision or an all-inclusive rent, the tenant will want the lease to include a clause requiring the landlord to provide and maintain the internet throughout their occupation of the premises. Even though the tenant is paying a sum to the landlord to cover the provision of "services", there is no implied obligation for the landlord to actually maintain the service.[52] As omnichannel retailing has caused retailers and their customers to become more reliant on the internet, it is important that tenants ensure that such a service is actually provided and maintained.

(b) Data collection

Data that is collected from landlord Wi-Fi connections is an area of contention for landlords and tenants operating in the digital age. There are two ways that customer data can be collected using customer Wi-Fi. Firstly, 'free Wi-Fi' systems can require users to set up a profile, usually requesting information such as date of birth or postcode. This information is highly desirable as it allows retailers and their landlords to understand their customer's demographic. Secondly, developments in technology have meant that Wi-Fi can be used to track the physical location of customers around shops and shopping centres. [53] This information can then be used to establish which areas of a shop are most and least popular.[54] From a data protection perspective, using Wi-Fi to track a customer's movement is far less intrusive than using technology like CCTV because no images of customers are retained and no personally identifying data is held.[55]

Since customer information is important to both landlord and tenant, it is advisable that there is a level of cooperation between the parties when negotiating the lease and it is agreed that the customer information collected by the customer Wi-Fi is shared between them.[56] It is in both of the parties' interests that the shop is targeting the right demographic and offering an enjoyable shopping experience. Depending on how technology advances, it may be quite straightforward for customer movement data collected by Wi-Fi to be analysed and separated between retailers, but it is likely that data consolidation and analysis will be an expensive and time-consuming task. From a landlord's perspective, the separation, analysis and sharing of Wi-Fi data is important as it can ensure that retailers who are operating within the shopping centre are targeting the correct demographic. Although data consolidation is expensive, a landlord of a multi-occupancy building, such as a shopping centre, is likely to be able to include the cost of analysing and sharing the data within the service charge. A sensible tenant would be unlikely to dispute this inclusion as the information provided to them would be of direct benefit, allowing them to better understand the shoppers who are entering their premises.

---

[51] P Pargma, "Commercial service charges: the occupier's perspective" (2004), available at *http://portal.cbre.eu/workspace/coming_soon/Commercial_Service_Charges_020204.pdf*.
[52] *Duke of Westminster v Guild* 1985 QB 688 at 697.
[53] Y Zeng and P Pathak, P Mohapatra, "Analyzing Shopper's Behavior through WiFi Signals" WPA '15 *Proceedings of the 2nd workshop on Workshop on Physical Analytics* (2015) 13.
[54] *Ibid*. 14.
[55] Landlords and tenants who use customer Wi-Fi to collect data must ensure that they comply with the Data Protection Act 1998.
[56] C Wells, "Clicks & Mortar: The marriage of the future and the past within a traditional retail lease structure - Part Two" (2016), available at http://www.shoosmiths.co.uk/client-resources/legal-updates/marriage-of-future-past-traditional-retail-lease-structure-2-11315.aspx.

(2) Turnover rent review

As explained in the previous section, retailers are prioritising the installation of Wi-Fi in stores to encourage omnichannel shopping. However, the fact that consumers can now make, and are encouraged to make, transactions in multiple ways has made 'turnover rent' an unworkable concept.

Turnover rent normally consists of two individual elements: a base rent and a 'top-up' amount based on the tenant's turnover.[57] The base rent is a fixed amount and usually equates to around 75% of the market value rent.[58] The top-up element tends to be a fixed percentage of the tenant's monthly turnover and this will fluctuate depending on the tenant's success.[59] The turnover rent system was imported from the United States in the 1990s as a more equitable way of calculating the level of rent that should be paid by a retail tenant.[60] It is considered to be a 'mutually beneficial rental system' as both parties end up having the same interest in the success of the retailer.[61]

Turnover rent provisions are dependent on the parties agreeing on a definition of 'turnover'. Prior to the advent of e-commerce, a retailer's turnover could only be linked to sales and returns made in a physical store, but integration of technology into traditional retail premises means that it is now possible for online orders to be researched, placed and returned in a physical store. The following sections will look at issues that may arise when the parties are trying to delineate the concept of turnover as a result of these changes in consumer behaviour and retail operations.

(a) Online orders made in store

Landlords will want online sales that are made in store by customers and staff to be included in the definition of turnover, but it is very difficult to establish whether an online order has been made in store or placed by a customer in the comfort of their own home.[62] In some shops, landlords have dealt with this issue by setting up an intranet system that enables shop assistants and customers to make online purchases in store and for these sales figures to be sent straight to the landlord.[63] Although such EPOS technology makes the appropriation of online sales data more straightforward, such a system may be expensive to design and maintain.[64] As leases are getting shorter, it is likely that a tenant will be unwilling to commit money to a system that a landlord may be able to use with future tenants. Even if a tenant held intellectual property rights over the system, it may still be aiming to reduce the size of its property portfolio or move its operations online and, as such, any investment in such a system would not be worthwhile.

If retailers were to invest in such a system, the lease would have to provide that it was properly maintained at all times. If the system is not properly maintained, a tenant could avoid paying the correct level of rent.[65] From a landlord's perspective, it may be preferable to carry out POS maintenance itself and pass the costs onto the tenant through a service charge. This way the landlord

---

[57] Gerber, *Commercial leases* para 9.04.
[58] A Howarth, "Turnover rents: Issues for landlords and tenants to consider before signing up" (2015), available at http://www.shoosmiths.co.uk/client-resources/legal-updates/issues-for-landlords-tenants-consider-before-signing-9364.aspx.
[59] Gerber, *Commercial leases* para 9.04.
[60] P Hendershot, "A comparison of upward only and turnover leases" (2002) 20(6) J of Property Investment & Finance 513 at 513.
[61] K Addae-Dapaah and C Yeo, "Percentage lease agreement as a shopping center management tool: a panacea for Singapore's retail industry woes?" (1999) 17(1) Property Management 24 at 26.
[62] C Barret and D Thomas, "Landlords fear web shopping fashion" (2011), available at https://www.ft.com/content/69e5157e-3871-11e0-959c-00144feabdc0.
[63] McAllister, *Scottish Law of Leases* 321.
[64] Electronic Point Of Sale (EPOS) systems are digital systems that allow customers to pay for goods electronically.
[65] P Clark, "Turnover rents: Part 2" (2014) 2 The Conveyancer and Property Lawyer 85 at 93.

can guarantee that the system is maintained properly, and the tenant is being charged the correct amount of rent.

(b) Online orders made at home

As well as customers making online orders in store, many choose to visit the physical store to conduct research and then order the product at home.[66] Landlords will want to include these sales in the definition of turnover because without the research conducted in the physical store it is possible that the consumer would not have purchased the product. Three possible solutions to this are proposed.

Firstly, some law firms suggest that landlords should negotiate a situation whereby the base rent is topped up by a percentage of the turnover generated by the store and then topped up further by a percentage of the turnover generated by the retailer's website.[67] This second percentage would account for products researched in store but subsequently ordered online. However, this model, which is currently being trialled by supermarkets operating in Australia, is unsatisfactory.[68] Determining what percentage of online turnover can be associated with a physical store is an impossible task. Although research has estimated that 31% of internet purchases made in the United Kingdom involve in-store research, the level of interaction with a physical store is highly dependent on the type of product sold by the retailer and its associated cost.[69] For example, if the customer is purchasing computer products, it is more likely that they will visit a physical store to conduct research.[70] In contrast, if the consumer is purchasing toys, books or DVDs from a retailer's website, they are less likely to conduct in-store research before making the purchase.[71] It is suggested that this difference in behaviour may be related to both the price discrepancy between these different categories of products and their relative life cycles. If a retailer sells only low-priced toys and books, it may be reasonable to assume that their consumers have had little interaction with a physical store. Conversely, if they sell expensive equipment it may be reasonable to assume that consumers have interacted with a physical store before making the online purchase. Given these issues, it does not appear that this is an acceptable solution.

Secondly, a 'catchment sale' concept could be attached to the definition of turnover rent to account for products researched in-store but subsequently purchased online. Turnover could be expanded to include any orders *delivered* within a particular radius of the shop.[72] However, as previously mentioned, online sales can be made without the customer carrying out in-store research. It is unfair to presume that a customer who has made an online purchase and had it delivered within a particular radius of a physical store has interacted with that store before making the online order.

Finally, it is possible that a retailer could integrate some level of enquiry into their online selling process to establish whether a customer visited a store before making a purchase. The inclusion of a radio button on the retailer's website could require the customer to confirm whether they visited a physical store before placing the online order.[73] A positive response would allow the money made from these sales to contribute to the tenant's in-store turnover. Unlike the other two

---

[66] Deloitte, "The omni channel opportunity: Unlocking the power of the connected consumer" (2014), available at *https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/consumer-business/unlocking-the-power-of-the-connected-consumer.pdf* .

[67] M Heighton, "Turnover rents- Some of the issues" (2002), available at *http://www.cms-lawnow.com/ealerts/2002/03/turnover-rent-leases8211-some-of-the-issues?cc_lang=en*.

[68] D Harington and K Dean, "Tenuous Turnover" (2015), available at *http://www.colliers.com/-/media/files/emea/easterneuropeaninformation/2014eeresearchpage/retail/ee-2015-tenuous-turnover-omni-channel-impact-final.pdf?la=en-gb*.

[69] Deloitte (n 66) at 8.

[70] G Browne, J Durrett and J Wetherbe, "Consumer reactions toward clicks and bricks: investigating buying behaviour on-line and at stores" (2004) 23(4) Behaviour and Information Technology 237 at 243.

[71] *Ibid.* at 243.

[72] M Heighton (n 67).

[73] A radio button is a small circle or box on a website that can be highlighted to make a choice.

solutions, this method does not involve estimating what percentage of sales can be attributed to a particular store. It would be an accurate way of gauging what sales were linked to a physical store, as long as the consumer was truthful when answering the question. However, similar to creating an intranet system, there would be significant costs associated with creating this online function, gathering and collating the information, and then transferring the information to the landlord. It is also possible that customers will purchase multiple products in one transaction and not all of them would be attributable to a visit to a physical store. The system would therefore have to ask customers about individual products and the burden that this would place on the customer may dissuade a retailer from adopting such a process.

Although they are all plausible, there is no infallible solution to expanding their turnover to include online sales made at home but researched in-store. Landlords must either accept that online sales will not be included in the calculation of turnover or tenants will have to accept that they could end up paying a level of rent that is not correlative to the turnover of their store. With demand for retail property low, it is likely that it will be the former, with landlords having to reach compromises to avoid vacancy.

(c) Online returns

Landlords will want to ensure that the in-store turnover of a tenant is not reduced by online refunds. If a customer purchases a product online and then returns it to a store, it is possible that the tenant could deduct the value of the product from the store turnover. One way of avoiding this manipulation is by introducing an inverse turnover rent structure.[74] This would require a fixed rent of market value to be agreed by the parties. If an agreed 'break point' was reached, there would be a percentage decrease in the rent paid by the tenant and this would discourage a tenant from using online returns to reduce store turnover. Although this model provides a possible solution, there remains a risk that a tenant would attempt to process in-store purchases as online returns in order to keep their store turnover high. In order for such a system to work, the landlord would have to ensure that the tenant's EPOS system was not capable of such avoidance.

(d) Footfall solution

Although commercial parties may prefer to continue using turnover rent in the digital age, the reality is that e-commerce has resulted in the delineation of 'turnover' becoming too complicated. There are difficulties if the sector continues to rely upon such a mechanism, despite challenges concerning online returns, online sales made in-store and online sales made because of in-store research.

It is therefore proposed that the commercial leasing sector should respond to the digital age by establishing a fair and balanced rent mechanism that is separate from the concept of turnover. As well as being difficult to define, turnover is no longer the sole aim of physical retail stores. Retailers now use physical stores to inspire customers to shop with them either online, via an app, in store or by telephone. Omnichannel retailing means that retailers are not concerned with whether customers make purchases when they are in store, provided they make a purchase using one of the available channels. As such, a better solution to turnover rent would be an inverse rent system based on the footfall that is achieved by a physical store.[75] The parties would agree upon a rent of market value and a break point. If footfall reached the break point, the tenant would pay a lower level of rent. An inverse rent review system would reward a tenant for increasing the footfall of its store and keeping the high street alive. Ultimately, it is in the landlord's interest to reward retailers who are keeping the high street alive, as their business depends on it.

---

[74] N Miller, "Retail leasing in a web-enabled world*"* (2000) J of Real Estate Portfolio Management 167 at 180. See 'Footfall solution' below.

[75] E Jones, "Click and capture? Harnessing multi-channel retail sales for turnover rent calculations" (2012), available at *http://www.lexology.com/library/detail.aspx?g=4cb837c2-9378-44b6-a7f6-b28f256f37eb*.

From an e-commerce perspective, there would be no benefit to a tenant trying to redirect online sales and refunds if rent was calculated by footfall. The landlord's concern would not be with the technicalities of a tenant's EPOS and ordering systems, but instead with ensuring that it was achieving a satisfactory level of footfall. Such a system would align the interests of the landlord and the tenant.[76]

Although such a system has benefits, there are some limits to its functionality. Subleasing part of a property is becoming more common in the digital age as retailers attempt to increase footfall into their shops.[77] This means that footfall figures may not be wholly attributable to the original tenant but to the subtenant. Therefore, a lease that relied upon a footfall level of rent would have to preclude subleasing or assignation of part of the property, a compromise that tenants may be unwilling to make in the digital age.

The main limit to the use of footfall rent is technology. Although footfall technology is used frequently by retailers to optimise store opening times, promotions and staffing levels, it has not been used to calculate rent.[78] It is suggested that the reason for this is footfall technology, which, although apt for establishing patterns and trends, is not accurate enough to be relied upon when calculating a variable rent. As technology develops, it is hoped that the commercial leasing sector will be able to respond to the digital age by creating a 'footfall' rent system that recognises the changing role of the physical store and the changing definition of 'retail success'.

D. CONCLUSION

The growth of e-commerce in Scotland has resulted in a fall in demand for retail property. Rather than offering low levels of rent, landlords can respond to and overcome high vacancy levels by being receptive to the retailer's demand for flexibility. Although a more cooperative and flexible approach to leasing may feel uncomfortable for landlords who have traditionally benefited from favourable lease terms, recognising the challenges faced by retailers will enable landlords to avoid vacant property. Whether flexibility is achieved by negotiating a qualified right to assign or sublet, more frequent break options, shorter lease durations or new pop-up lease types, the cumulative effect of these responses will be that landlords are less likely to be left with unprofitable vacant property. If landlords take a flexible approach to lease negotiation this may encourage retailers to continue to occupy physical premises and the demise of the high street may be avoided.

Additionally, the adoption of omnichannel retailing strategies has meant that retailers' in-store operations have changed. They now integrate significantly more technology, and this has created new areas of consideration when entering a lease. In particular, customer Wi-Fi is more popular in retail premises and, as such, the introduction of data collection is likely to cause some controversy. The main impact of this increased reliance on internet connectivity is that turnover rent is no longer a viable mechanism for determining rent. It is suggested that adjusting the concept of turnover is too complicated a task, resulting in either a disappointed retailer or a disappointed landlord. As such, the Scottish commercial leasing sector should respond with a flexible rent provision that is based on something other than turnover. In this article, it is suggested that a rent review provision based on footfall could be a potential solution to the inadequacy of turnover rent. However, for such a solution to be workable and successful, it is likely that further technological development will be required.

[76] L Gallosti, "Footfall analytics: insights for smart retail leasing" (2016), available at *http://www.wsp-pb.com/en/What-we-do/Strategic-Consulting/Insights/Footfall-analytics-insights-for-smart-retail-leasing/*.

[77] 68% of the United Kingdom's top retailers are attempting to share their premises with a leisure operator in order to drive footfall into their store. See TLT Solicitors (n 2) at 12.

[78] M Kirkup, "Electronic footfall monitoring: experiences among UK clothing multiples" (1999) 27(4) International Journal of Retail & Distribution Management 166 at 167.

That said, technology is developing rapidly, and it is possible that such a mechanism will be established in the near future.

EUROPEAN UNION CITIZENSHIP: AN ASSESSMENT OF THE PURPOSE AND SUBSTANCE OF EU CITIZENSHIP AS AN AUTONOMOUS STATUS

Maya Moss[*]

A. INTRODUCTION

As a supranational entity, the European Union (EU) is unique in its powers and scope, having developed from a purely economic Union, to something far more encompassing, and claiming to be a 'European citizen' is now a commonplace statement. Extensively debated in academic literature, its hold on the public has not lessened or diminished, perhaps also clearly evidenced by the debates in the UK regarding whether, and if so how, British citizens might hold onto their European citizenship and rights after next year's exit. Its existence is dependent upon an individual holding national citizenship in a member state, and many rights gained as a result of Union citizenship are, at least traditionally, dependent on cross-border movement to bring the situation within the sphere of EU law. This article seeks to question the purpose of Union citizenship when it has these ties, focusing specifically on the ties to free movement, considering how useful European citizenship can truly be where a person must establish some cross-border element in order to benefit from many of the rights therein. In cases where economically inactive citizens may find protection, there exists a lack of clarity in the consistency of the Court of Justice of the European Union's (CJEU) approach after the judgment in *Dano.*[1] The possibility of such a line of cases would undermine previous efforts to expand the scope of protection for non-worker citizens. It is argued, in this vein, that there remains an economic focus within the EU as regards citizenship.

[*] LLM Student in European Law at the University of Edinburgh.
[1] Case C-333/13 *Dano v Jobcenter Leipzig*, 11 November 2014.

The first part deals with Union citizenship generally, outlining both its legislative underpinning and the Court's development of its approach through some of the case law. Its aim is to demonstrate that Union citizenship may have broadened the previous economic approach of the Court and the EU, but that certain inconsistences in the approach of the Court are problematic for the coherence of citizenship. The second part explores the relationship between free movement and citizenship, and the way in which movement has become less necessary than a cross-border element. Despite the Court's push for generosity, it is argued that the extent of protection stemming from citizenship remains limited. In all cases there remains some link to EU law, and although it is not the contention of this article that no such link should exist, it systematically excludes static citizens, who are by virtue of the Treaty still citizens. The final part discusses further ramifications of EU citizenship, particularly in relation to state-based national citizenship. Given the difficulty of the notion of identity as regards Union citizenship, it is argued that Union citizenship defies much of what is expected of a status such as citizenship, and even in attempts to mitigate this – such as by enfranchising Union citizens in European and municipal elections – the problem remains in that Union citizens are still not treated equally. Apart from the specific exceptions set by the *Ruiz Zambrano* line of cases, as well as the prisoner voting principle in *Delvigne,*[2] both of which are specific on the facts, Union citizenship remains tied to cross-border situations and thus struggles to outweigh the benefits of national citizenship.

B. EUROPEAN UNION CITIZENSHIP

(1) Union citizenship in legislation

EU citizenship was introduced by the Maastricht Treaty in 1992,[3] currently found in art 20 of the Treaty on the Functioning of the European Union (TFEU).[4] It is additional to, and does not replace, national citizenship,[5] thereby "dependent on national belonging".[6] Art 21 TFEU further extends the right to move and reside freely within the Union to all citizens of the EU. The Court used to be primarily concerned with the free movement and rights of European workers, goods, and capital, known generally as 'market citizenship'.[7] This economically focused attitude that informed the early free movement of workers/persons provisions may have lessened with the introduction of Union citizenship, the objective of which was to encourage a European identity and define the relationship between people and the EU, through "rights, duties and political participation," in order to create a political consciousness throughout Europe.[8]

(2) From worker to citizen

The Treaty provides for free movement within the Union,[9] abolition of discrimination in employment[10] and the right to accept offers for employment, move within the territory of member state for that purpose, to stay in a member state for the purpose of employment, and to remain in that territory after having been employed.[11] It has been argued that the rationale behind the free movement

[2] Case C-650/13 *Delvigne v Commune de Lesparre Medoc and Prefet de la Gironde*, 6 October 2015.
[3] Art 8 of the Treaty on European Union (TEU).
[4] The consolidated version of the TFEU published at OJ C326/47 2012.
[5] Art 20(1) TFEU.
[6] T Konstadinides, "La fraternite europeene? The extent of national competence to condition the acquisition and loss of nationality from the perspective of EU citizenship" (2010) 35(3) EL Rev 401.
[7] D Kostakopoulou, "Ideas, norms and European Citizenship: Explaining institutional change" (2005) 68(2) MLR 233 at 234.
[8] European Parliament, "Factsheet: the citizens of the Union and their rights", available at *http://www.europarl.europa.eu/facts_2004/2_2_0_en.htm*.
[9] Art 45(1) TFEU.
[10] Art 45(2) TFEU.
[11] Art 45(3) TFEU.

provisions was to "stimulate labour movements" to regions where labour demand was higher, thereby improving working and living conditions throughout.[12] Some contend that the right to free movement, even in light of European citizenship, is primarily aimed at economically active individuals.[13] Such an assertion might be supported by the hierarchy found in Directive 2004/38 on citizenship rights, wherein right of residence for longer than three months is granted firstly to all Union citizens if they are workers, and secondly, if they have sufficient resources.[14]

The early case law informed by the notion of European citizenship was generally accommodating. In *Martinez Sala*,[15] the Court held that requiring a non-national to provide a residence permit where a similar condition was not placed on nationals amounted to unequal treatment.[16] By virtue of her status as a Union citizen lawfully resident in another member state, she could rely on the Treaty in any situation within the scope of EU law.[17] The limits of the reach of the CJEU, however, may be illustrated by *Wijsenbeek*,[18] in which the Court clarified that passport control when crossing internal borders was not an obstacle to free movement, nor was it disproportionate.[19] The Court highlighted that with a lack of harmonising of laws regarding crossing of internal borders, immigration, visas, and asylum, finding border controls contrary to EU law was impossible.[20] Dora Kostakopoulou has argued that the Court in *Sala* offered a "new constructive meaning" of citizenship,[21] illustrating its creativity, given that there was no universal agreement on the exact status of non-worker citizens.[22] However, as seen in *Wijsenbeek*, it is also clear that the Court would not take on an overly creative role, in disagreement with the Commission,[23] remaining cautious in the interpretation of free movement for Union citizens without an economic element.

(a) The Citizenship Directive

The 'Citizenship Directive' 2004/38 (henceforth 'the Directive') applies to "all Union citizens who move to or reside in a member state other than that of which they are a national", as well as their family members.[24] Beyond three months, there are remnants of the economic activity necessity from pre-Union citizenship, requiring that the citizen is a worker or self-employed, or that they have sufficient resources and comprehensive sickness insurance.[25] In certain cases, Union citizenship appears also to have been used as a means of furthering economic aims. In *Collins*, the Court held that it was not possible, post-Union citizenship, to exclude from the scope of equal treatment benefits intended to facilitate access to the employment market.[26] Although the language attempts to distance itself from the economic activity necessity, there remains an admission that member states may limit prolonged residence in order to protect the national welfare system. For non-worker citizens, the economic element is still present in the form of ensuring that the rights holder does not become an "unreasonable burden" on the national welfare system.[27] Workers are generally more extensively protected than economically inactive Union citizens, where the term "worker" includes job-seekers

[12] S O'Leary, "The Past, Present and Future of the Purely Internal Rule in EU Law" in M Dougan, N Nic Shuibhne and E Spaventa (eds), *Empowerment and Disempowerment of the European Citizen* (2012) 37 at 41.
[13] H Verschueren, "Free Movement of EU Citizens Including for the Poor?" (2015) 22(1) Maastricht Journal of European and Comparative Law 10 at 12.
[14] European Parliament and Council Directive 2004/38/EC OJ 2004 L158/77 art 7.
[15] Case C-85/96 *Martinez Sala v Freistaat Bayern* [1998] ECR I-02691.
[16] *Ibid.* at para 54.
[17] *Ibid.* at para 63.
[18] Case C-378/97 *Criminal proceedings against Florus Ariël Wijsenbeek* [1999] ECR I-06207
[19] *Ibid*. at para 36.
[20] *Ibid.* at para 40.
[21] Kostakopoulou (n 7) at 247.
[22] D Thym, "When Union citizens turn into illegal migrants: the Dano case" (2015) 40(2) EL Rev 249 at 253.
[23] Case C-378/97 *Criminal proceedings against Florus Ariël Wijsenbeek* [1999] ECR I-06207 at para 35.
[24] European Parliament and Council Directive 2004/38/EC OJ 2004 L158/77 art 3.
[25] *Ibid.* art 7(1)(a) and (b).
[26] Case C-138/02 *Collins v Secretary of State for Work and Pensions* [2004] ECR I-02703 at para 63.
[27] European Parliament and Council Directive 2004/38/EC OJ 2004 L158/77Recital 10.

even after a period of six months, allowing them to remain while they are still searching for, and have a "genuine chance" of finding, work.[28] It might also include those in social reintegration programmes, as in *Trojani*.[29]

(b) Growing importance

In *Grzelczyk*,[30] the Court espoused a principle that Union citizenship was to become the "fundamental status" of all those who held it,[31] now in Recital 3 of the Directive. The Court also stressed that there existed some degree of "financial solidarity" between nationals and non-nationals in various member states.[32] Another key case in the development of Union citizenship and its scope is *Baumbast*.[33] Here, the issue was of a Union citizen from Germany, married to a third-country national,[34] seeking residence permits for indefinite leave in the UK.[35] The couple had a daughter with dual German and Colombian nationality,[36] were not receiving social benefits, and had "comprehensive medical insurance" in Germany.[37] The Court highlighted that preventing a child of a Union citizen from continuing their education in the host member state "might dissuade that citizen from exercising the rights to… movement," constituting a hindrance to the effective exercise of that freedom.[38] Important in this respect was Regulation No 1612/68, allowing access to educational courses for children of Union citizens who are or have been employed in another member state; not allowing a primary carer to remain with the child would infringe that right.[39] Importantly, the Court confirmed that the rights of citizenship on entry and residence in the Treaty were directly effective,[40] and appeared to separate citizenship from its previous market attachment by holding that even citizens who had ceased to be migrant workers could remain in order to facilitate the right of children in education.[41] In line with the objective of Union citizenship, these decisions may be viewed in light of the desire to create a European identity,[42] by making residence more readily available.

(3) The purely internal rule

In the case of *Saunders*, the Court denied that provision on free movement of workers could be applied without connection to "situations envisaged by Community law."[43] However, in the citizenship case law, it has become clear from a number of cases that such situations may arise even where no movement has occurred. *Garcia Avello*[44] concerned two children with dual Spanish and Belgian citizenship.[45] The Court held that a link was established with Community law where nationals of one member state resided in another,[46] and Union law therefore precluded member states from

[28] Case C-292/89 *The Queen v Immigration Appeal Tribunal ex parte Antonissen* [1991] ECR I-00745 at para 21.
[29] Case C-456/02 *Trojani v Centre Publique d'aide sociale de Bruxelles* [2004] ECR I-07573 at para 22.
[30] Case C-184/99 *Grzelczyk v Centre public d'aide sociale d'Ottignies-Louvain-la-Neuve* [2001] ECR I-06193.
[31] *Ibid*. at para 31.
[32] *Ibid*. at para 44.
[33] Case C-413/99 *Baumbast and R v Secretary of State for the Home Department* [2002] ECR I-07091.
[34] *Ibid*. at para 16.
[35] *Ibid*. at para 20.
[36] *Ibid*. at para 16.
[37] *Ibid*. at para 19.
[38] *Ibid*. at para 52
[39] *Ibid*. at para 73.
[40] *Ibid*. at para 84.
[41] *Ibid*. at para 75.
[42] European Parliament Factsheet (n 8).
[43] Case 175/78 *The Queen v Saunders* [1979] ECR 01129 at para 11.
[44] Case C-148/02 *Garcia Avello v Belgian State* [2003] ECR I-11613
[45] *Ibid*. at para 13.
[46] *Ibid*. at para 27.

refusing a change of surname where the person in question had dual nationality of that member state and another, where the other member state entitled the person to such a surname.[47]

Ruiz Zambrano[48] concerned a Colombian couple applying for refugee status in Belgium.[49] The application was refused, but as a result of the non-refoulement clause they could not be sent back to Colombia.[50] Mr Zambrano signed a contract for work despite not having a valid work permit,[51] and the couple later had another child who gained Belgian nationality.[52] All governments that submitted observations contended that the situation did not fall within the scope of EU law, as the children were nationals of the state in which they resided.[53] The Court, however, held that art 20 TFEU prevented member states from applying measures that had the "effect of depriving citizens of the Union of the genuine enjoyment of the substance of the rights" granted by the Union.[54] As such, refusing to grant residence to a third-country national with minor children in the member state of nationality of those children had such an effect.[55]

The Court later reduced protection in similar cases. One example is *Alokpa*,[56] concerning a Togo national who was at first denied international protection in Luxembourg,[57] but later granted discretionary leave to remain upon the birth of twins,[58] who gained French nationality by virtue of their father.[59] The applicant had no relationship with the father, and applied for a residence permit in order to ensure appropriate care needed for the children. However, such a permit was also rejected.[60] The Court held that there was an "intrinsic connection" between the free movement rights of Union citizens, and the ability of third-country nationals to enter and reside in the Union so as to not interfere with that freedom.[61] The Court went on to hold that the mother could have applied for derived residence in France[62] and that it was up to the national court to decide whether removal of the mother would reduce the children's enjoyment of their rights. The Court paid little attention to the fact that there was no practical connection with France, other than the legal factor of nationality. Such an artificial interpretation stands in stark contrast to the stance taken in *Zambrano*, and might indicate a degree of hesitancy from the Court's side.

(4) *Dano:* a limitation?

The Court's generous application of the citizenship provisions was contradicted in *Dano*.[63] The applicant, a Romanian national residing in Germany, lived with her son in her sister's flat. Her sister provided for them.[64] Ms Dano had been residing in Germany for more than three months, was not searching for employment, and she had not entered Germany for work. As a result, she fell outside the personal scope of Directive 2004/38.[65] The Court continued by stating that accepting that those

[47] *Ibid*. at para 45.
[48] Case C-34/09 *Ruiz Zambrano v Office national de l'emploi* [2011] I-01177.
[49] *Ibid*. at para 14.
[50] *Ibid*. at para 15.
[51] *Ibid*. at para 18.
[52] *Ibid*. at para 19.
[53] *Ibid.* at para 37.
[54] *Ibid*. at para 42.
[55] *Ibid*. at para 43.
[56] Case C-86/12 *Alokpa and Moudoulou v Ministre du Travail, de l'Emploi et de l'Immigration,* 10 October 2013.
[57] *Ibid*. at para 11.
[58] *Ibid*. at para 12.
[59] *Ibid*. at para 13.
[60] *Ibid*. at para 15.
[61] *Ibid*. at para 23.
[62] *Ibid*. at para 34.
[63] Case C-333/13 *Dano v Jobcenter Leipzig*, 11 November 2014.
[64] *Ibid*. at para 37.
[65] *Ibid*. at para 66.

outside the scope of the Directive could claim social benefits "under the same conditions" as nationals would undermine the objective of avoiding Union citizens becoming an 'unreasonable burden'.[66] Resources were narrowly interpreted, despite the earlier decision that it should include "available" resources, including provision by, for example, a third-country national,[67] supported recently in *NA*.[68] This notion of availability is clearly broader than assessing the actual funds of a Union citizen. The CJEU ultimately stated that she could legitimately be refused benefits because she fell outside the Directive's scope.[69]

The Court in *Dano* seemed most concerned about Ms Dano's perceived lack of connection to the host state, stepping back from a previously creative role in what has been called "doctrinal conservatism".[70] The Court later held that a single applicant cannot be considered an "unreasonable burden".[71] Ms Dano's circumstances did not seem to indicate a high likelihood of her becoming an "unreasonable burden", seemingly in line with the test suggested in *Brey*,[72] making an assessment of the "specific burden" that the granting of the benefit would have on the national assistance system "as a whole".[73] As such, it is not surprising that the exact meaning of this phrase remains unclear,[74] and the restrictive attitude taken in *Dano* betrays both a political concern,[75] and an economic one, given the legitimate wish to protect national welfare systems from abuse.[76]

The Court's finding contradicts that in *Brey*, in that it seems to allow member state to disregard benefits that enable a person to meet the sufficient resources requirement.[77] The suggestion is that the divide between economically active and inactive "free movers has been petrified" because the Court still allows for unequal treatment.[78] *Dano* indicates a situation in which unequal treatment is permitted, despite a goal of citizenship surely being to ensure equal treatment to all,[79] particularly if it is to remain at all meaningful.[80] In a strict interpretation of *Dano*, one might hold that the Court simply limited the scope of the derogation to the equal treatment principle to situations where the "only motive for moving… is to obtain social assistance".[81] Equally, the Court did not appear to limit the derogation to social assistance benefits, but refers broadly to the national "welfare system", without reference to any proportionality test in this regard.[82] The inconsistencies in the Court's jurisprudence, an example being the sudden acceptance of the *Zambrano* judgment, might indicate a possibility that *Dano* returns when required by political or other perspectives.

The endorsements of *Zambrano* in more recent case law[83] indicate a broad reading of citizenship, considering the cross-border element practically rather than literally, meaning that a

---

[66] *Ibid*. at para 74.
[67] Case C-86/12 *Alokpa and Moudoulou v Ministre du Travail, de l'Emploi et de l'Immigration,* 10 October 2013 at para 27.
[68] Case C-115/15 *Secretary of State for the Home Department v NA*, 30 June 2016 at para 77.
[69] Case C-333/13 *Dano v Jobcenter Leipzig*, 11 November 2014 at para 84.
[70] Thym (n 22) at 250.
[71] Case C-67/14 *Jobcenter Berlin Neukolln v Alimanovic,* 15 September 2015 at para 62.
[72] Case C-140/12 *Pensionsversicherungsanstalt v Peter Brey,* 19 September 2013.
[73] *Ibid.* at para 64.
[74] S de Mars, "Economically inactive EU migrants and the United Kingdom's National Health Service: unreasonable burdens without real links?"(2014) 39(6) ELR 770 at 771.
[75] Thym (n 2222) at 253.
[76] Verschueren (n 13) at 24.
[77] D Düsterhaus, "Timeo Danones et done petentes" (2015) 11(1) ECL Rev 121 at 134.
[78] *Ibid.* at 138.
[79] *Ibid.*
[80] E Spaventa, "Earned citizenship – understanding Union citizenship through its scope", in D Kochenov (ed), *EU Citizenship and Federalism: the Role of Rights* (2015) at 1.
[81] Verschueren (n 13) at 26.
[82] *Ibid*. at 28.
[83] Case C-165/14 *Alfredo Rendón Marín v Administración del Estado* 13 September 2016; Case C-304/14 *Secretary of State for the Home Department v CS,* 13 September 2016; Case C-115/15 *Secretary of State for the Home Department v NA,* 30 June 2016.

connection with more than one member state, where a national measure may have the effect of weakening Union protection, is sufficient. Further, these illustrate a continued decoupling of citizenship from ties to other provisions in an attempt to protect as broad a range of people as feasible. The relationship between Union citizenship and movement will be further explored in the next section, but it is highlighted here to emphasise the inherent limitations in the further development of the Union citizenship status.

C. THE RELATIONSHIP BETWEEN CITIZENSHIP AND MOVEMENT

(1) Free movement and Union citizenship

The triggering link in the Directive is necessary by virtue of the principle of conferral,[84] and in the context of early free movement case law, the purely internal rule was justified in order to include in the scope of EU law only issues that were "sufficiently connected with the (economic) aims underpinning" the free movement provisions.[85] Without this, there exists no reasonable connection to the EU or its legislation. Residing in a member state other than that of nationality, however, does not always necessitate physical movement. It is clear, therefore, that the provision does not make actual movement necessary, but rather that it requires a connection with more than one member state; in this regard, the use of 'cross-border' has become somewhat misleading. Some have contended that this focus on crossing borders contradicts EU law on the internal market, in that the citizenship discussion has made "indispensable" the discussion of boundaries, which the internal market seeks to remove.[86] In this regard, whether or not actual movement is necessary does not matter. The fact remains that the language and approach of the Court is based upon the existence of such borders.

(2) Union citizenship without movement

Neither the preamble nor any of the citizenship provisions in the Treaties require economic activity or movement to make use of citizenship.[87] The argument in this article is that the necessity of a 'European' element limits the reach of Union citizenship despite the Court's efforts to expand it. Paradoxically, removal of this element and the reduction of the scope of 'internal' would surely take away from the very 'Europeanness' that ought to inform Union citizenship. Despite this, some claim that member state nationals "enjoy substantive citizenship rights… independently of movement," by referring to the decision in *Zambrano*.[88]

Another noteworthy case is that of *Rottman*.[89] This case concerned an Austrian national,[90] who upon applying for German nationality did not disclose pending criminal proceedings against him in Austria, and German naturalisation was granted.[91] In line with Austrian legislation, his naturalisation in Germany meant that he lost his Austrian nationality.[92] When German authorities were notified of the proceedings,[93] naturalisation was withdrawn.[94] The Court held that it was not

[84] Art 5(2) TEU.
[85] O'Leary (n 12) at 39, emphasis added.
[86] D Kochenov and R Plender, "EU citizenship: from an incipient form to an incipient substance? The discovery of the treaty text" (2012) 37(4) EL Rev 369 at 384.
[87] *Ibid*. at 377.
[88] A Wiesbrock, "Disentangling the "Union citizenship puzzle"? The McCarthy Case" (2011) 36(6) EL Rev 861 at 867.
[89] Case C-135/08 *Rottman v Freistaat Bayern* [2010] ELR I-01449.
[90] *Ibid*. at para 22.
[91] *Ibid*. at para 25.
[92] *Ibid*. at para 26.
[93] *Ibid*. at para 27.
[94] *Ibid.* at para 28.

contrary to EU law to withdraw member state nationality obtained by deception, provided it was proportionate.[95]

The German and Austrian governments contended that the situation was purely internal,[96] and several other member states believed it fell outside the scope of the Treaty.[97] The Court emphasised that because withdrawal of his naturalisation could lead to him losing his status as a Union citizen,[98] such a measure would have to be proportionate;[99] a necessity that had not previously been (explicitly) required. The Court held that even in cases falling *prima facie* outside the scope of the Treaty, national courts must have "due regard" for EU law,[100] and thus, where there was a risk of losing Union citizenship status, the case must fall within the scope of the Treaty.[101] With the added result of statelessness, the consequences of withdrawing naturalisation in this case were arguably more adverse than those in *Zambrano*, given that while the Zambrano children's citizenship would have had little practical effect if they were prevented from residing within the territory of the Union, Rottman's complete loss of that status would make impossible any present or future exercise of those rights. The Court could, however, have made a point of the prior movement in order to maintain consistency in the relationship between movement and citizenship. This would not necessarily have detracted from the contention that the loss of EU citizenship would also be enough to bring an issue within the scope of EU law.

In *McCarthy*,[102] it became clear that the protection in *Zambrano* was especially narrow. McCarthy, a citizen with dual British and Irish nationality, having always resided in the UK,[103] could not gain such protection for her Jamaican spouse.[104] The Court held that while a lack of movement did not automatically make a situation purely internal,[105] nothing in that case had "the effect of depriving her of the genuine enjoyment of the substance of [her] rights."[106]

The furthering of the *Zambrano* protection in recent case law, however, has continued to emphasise the Court's willingness to extend protection to a broader range of people. The Court highlighted in *CS* that the purpose of the derived rights for third-country national family members was to avoid interfering with the Union citizen's right to free movement.[107] There exists, therefore, a possibility that Union citizenship may grant wide protection even in cases that are, at first glance, purely internal. It is contended, however, that these cases, just as *Zambrano,* are specific. This was recognised by the Court itself in *NA*, where it held that the issue of "deprivation of the substance of [citizenship] rights" was specific to situations relating to residence of third-country nationals who are family members of dependent Union citizens.[108] This indicates that such protection could not apply to, for example, citizen parents who do not fulfil the requirements of the Directive, such as a situation comparable to that in *Dano*. Such parents would clearly always have residence in their member state of nationality, so even if children of such parents had a different nationality, they would not be made to leave the Union.

[95] *Ibid.* at para 59.
[96] *Ibid.* at para 38.
[97] *Ibid*. at para 37.
[98] *Ibid*. at para 56.
[99] *Ibid*. at para 59.
[100] *Ibid*. at para 41.
[101] *Ibid*. at para 42.
[102] Case C-434/09 *McCarthy v Secretary of State for the Home Department* [2011] ECR I-03375.
[103] *Ibid.* at para 14.
[104] *Ibid*. at para 17.
[105] *Ibid*. at para 46.
[106] *Ibid*. at para 49.
[107] Case C-304/14 *Secretary of State for the Home Department v CS,* 13 September 2016 at para 28.
[108] Case C-115/15 *Secretary of State for the Home Department v NA,* 30 June 2016 at para 72.

The finding of a cross-border element in cases where a person's EU citizenship is at risk has sometimes appeared strained, or "artificial."[109] The presumed "disconnection" between cross-border elements as triggering EU law and the objectives of EU citizenship has "undermined the legitimacy and predictability" of the Court's jurisprudence.[110] In other words, the case law does not necessarily make clear what constitutes a cross-border element, and in what circumstances other connecting factors may be permissible. It is the argument of this article that, at present, such a cross-border element must matter in order for the Directive to remain useful and sensible. Indeed, if citizenship itself were a sufficient link with Union law, several provisions of the Directive would need amending.[111] Additionally, as the purpose of the purely internal rule is to protect areas of national competence, simultaneously calling for its removal is somewhat "incoherent."[112] This clearly illustrates the dilemma faced by the Court in its further development of Union citizenship, and it is one that cannot merely be disregarded, particularly given the politically sensitive nature of division of competences.

(3) Legal uncertainty

Continuing to force a finding of cross-border elements merely creates inconsistencies, which undermine the importance of Union citizenship, given that free movement and residence is only one right of many conferred upon Union citizens.[113] On the one hand, the qualifications of "enjoyment of the substance" of citizenship rights, as seen above, limit which internal situations might fall within the scope of the Treaty, but it is formulated in an open manner, likely requiring determination on a "case-by-case basis."[114] This causes further uncertainty in whether citizenship would be relevant in any given case. The Court has later (*NA, CS, Marín*) been more specific in terms of the "genuine enjoyment" test from *Zambrano*, and in that regard it would appear that the rule or principle emerging from these cases is specific to the type of case concerned. The Court in *NA* states that the "substance of the rights" concerned is specific to the case of a third-country national family member of a citizen, who would be deprived of the genuine enjoyment of those rights as a result of the national measure.[115] It would seem, then, that under other circumstances, this test of "substance" and "genuine enjoyment" would not apply at all. It is clear that the threshold of "genuine enjoyment" is high,[116] and thus unlikely to apply to circumstances that are less unfavourable. If this uncertainty persists, the fact that these cases are specific does not alleviate the aforementioned concern as to whether the cross-border element remains necessary in cases that are not similar to these.

(4) Reverse discrimination

Reverse discrimination occurs when national legislation offers less protection than European legislation, placing those outside the scope of EU law at a disadvantage. The main reason for this problem is that many citizenship rights are guaranteed through a "model of free movement."[117] The persistence of this problem does not appear to recognise the "qualitative change" in the aims of the EU post-citizenship, from economic to social and political, which continues to be seen most clearly in

[109] S Reynolds, "Exploring the "intrinsic connection" between free movement and the genuine enjoyment test: reflections on EU citizenship after Iida" (2013) 38(3) EL Rev 376 at 377.
[110] P Van Elsuwege, "European Union citizenship and the purely internal rule revisited: Decision of 5 May 2011, Case C-434/09 Shirley McCarthy v. Secretary of State for the Home Department" (2011) 7(2) ECL Rev 308 at 316.
[111] C Dautricourt and S Thomas, "Reverse discrimination and free movement of persons under Community law: all for Ulysses, nothing for Penelope?" (2009) 34(3) EL Rev 433 at 450.
[112] O'Leary (n 12) at 64.
[113] *Ibid.* at 47.
[114] S Mantu, "European Union citizenship anno 2011: Zambrano, McCarthy and Dereci" [2012] 26(1) Journal of Immigration Asylum and Nationality Law 40 at 52.
[115] Case C-115/15 *Secretary of State for the Home Department v NA,* 30 June 2016 at para 72.
[116] Elsuwege (n 110) at 314.
[117] H Van Eijken and S A de Vries, "A new route into the promised land? Being a European citizen after Ruiz Zambrano" (2011) 36(5) ELR 704 at 710.

the context of reverse discrimination.[118] Moreover, this makes it inevitable that the view that European law protects only a "privileged minority of its citizens" is reinforced.[119] Given the resources needed to move between member states, many of those who do are not likely to be considered in need of assistance.[120]

One example of the possibility of a wide reading of reverse discrimination may be found in *Carpenter*.[121] Mrs Carpenter, a third-country national married to a UK national, overstayed her leave to remain in the United Kingdom prior to her marriage.[122] Mrs Carpenter argued that she had a right to remain because deporting her would hinder her husband's "right to provide and receive services" as she took care of his children during Mr Carpenter's travels.[123] The Court considered that the freedom to provide and receive services "could not be fully effective" if he was hindered in his exercise thereof "by obstacles raised in his country of origin to the entry and residence of his spouse,"[124] thus deporting Mrs Carpenter would "not strike a fair balance."[125]

If this situation had arisen purely within the UK, it would not have come within the scope of EU law on family reunification. There is a "weak economic component and tenuous cross-border element,"[126] making the connection between Mrs Carpenter's residence and the provision of services far-fetched,[127] particularly since deporting Mrs Carpenter would not have wholly deprived Mr Carpenter of the ability of perform the relevant services. The Court exaggerated the detrimental impact Mrs Carpenter's deportation would have on the ability of Mr Carpenter to provide services. From the perspective of coherence this is questionable, especially how it fits in with other free movement case law. The more recent case of *O and B*[128] illustrates more specifically how cases of Union citizens moving between their member state of nationality and a host member state might connect them to citizenship provisions. In particular the Directive may be applied by "analogy"[129] in circular cases where a citizen returns to their member state of nationality after having "created or strengthened" a family life in a host member state.[130]

Willem Maas has argued that the introduction of citizenship ought to "reduce and ultimately eliminate reverse discrimination" because it alters the relationship between citizens and their governments, and offers equality of treatment.[131] It is submitted that the existence of reverse discrimination itself undermines this equality of treatment amongst a specific group of people, namely those who have not moved, and cannot find any connecting factor, usually "novel",[132] to link their situation to EU law. As such, it might be argued that reverse discrimination falls within the non-discrimination provisions of the Treaty,[133] which would mean that the provisions on non-discrimination contradict those in the Directive limiting its application to cross-border situations. This

[118] O'Leary (n 12) at 61.
[119] Dautricourt and Thomas (n 111) at 436.
[120] A José Menéndez, "European Citizenship after Martínez Sala and Baumbast: Has European law become more human but less social?" Working Paper No.11 June 2009, ARENA Working Paper, available at *http://www.sv.uio.no/arena/english/research/publications/arena-working-papers/2001-2010/2009/WP11_09_Online.pdf* at 40.
[121] Case C-60/00 *Mary Carpenter v Secretary of State for the Home Department* [2002] ELR I-06279.
[122] *Ibid*. at para 13.
[123] *Ibid*. at para 17.
[124] *Ibid.* at para 39.
[125] *Ibid*. at para 43.
[126] O'Leary (n 12) at 53.
[127] Elsuwege (n 110) at 316.
[128] Case C-456/12 *O and B*, 12 March 2014.
[129] *Ibid*. at para 50.
[130] *Ibid*. at para 49.
[131] W Maas, "The Origins, Evolution, and Political Objectives of European Citizenship" [2014] 15(5) German Law Journal 797 at 802.
[132] Reynolds (n 109) at 377.
[133] O'Leary (n 12) at 61.

is problematic, as citizenship should not be taken to expand the scope of the EU's powers generally.[134] But an interpretation that discrimination against static citizens falls within the scope of the Treaty would surely have such an effect. Such a broad interpretation in the context of, for example, fundamental rights may also have that effect, risking affecting the regulatory autonomy of member states as protected by the TEU.[135] Equally, the inconsistency may remain were static citizens able to rely on citizenship provisions, in that citizenship of the Union is explicitly "complementary" in nature, and holding otherwise would have "pervasive constitutional consequences."[136]

On the other hand, one might view reverse discrimination purely as discrimination against those who have not moved, having chosen not to make use of an EU right,[137] or, in the context of the internal market, have not contributed thereto.[138] This is clearly a semantic issue, but it serves to illustrate the particular group that is being discriminated against from an EU perspective. These people are placed at a disadvantage by their decision not to move. It is clear that the EU cannot and should not intervene where it has not been conferred powers. To contend otherwise would be to disregard the EU's self-imposed constitutional restrictions, and arguably to take judicial decision-making too far. Reverse discrimination illustrates a clear boundary of EU influence which serves to undermine the strength of protection otherwise offered to EU citizens.

D. RAMIFICATIONS AND CREDIBILITY OF UNION CITIZENSHIP

(1) Ramifications

A definition of citizenship tends to point to a "collection of rights and obligations" resulting in a "formal juridical identity" for those who hold it.[139] As much as it denotes who is part of a certain community, citizenship is a "powerful [instrument] of exclusion."[140] Jo Shaw argued in the late 1990s that Union citizenship must enable those in the EU to "identify each other within the political processes of European integration…"[141] and that given the complex nature of citizenship it "cannot simply be transplanted" between political and economic settings.[142] Union citizenship tends to be viewed in the same way as national citizenship, or at least in the same category of statuses a person might hold, especially as a result of the fact that it is the nationality of individual member states which determines who is a Union citizen. The relationship between Union and national citizenship is rather a "matter of convenience," and does not as such tie together all those who hold the status.[143] Arguably there is no alternative way of conceiving citizenship with reference to real examples. Given that there does not exist a "European *demos* in the ethnic, psycho-sociological, cultural/civilisational or emotional" sense,[144] the notion of European citizenship rests very much upon the strength of the legal rights associated with it. As such, it is clearly problematic to conceive of Union citizenship as equivalent or comparable to national citizenship.

---

[134] Case C-212/06 *Government of the French Community v Flemish Government* [2008] ELR I-01683 at para 39.
[135] Elsuwege (n 110) at 322.
[136] E Spaventa, "Seeing the wood despite the trees? On the scope of Union citizenship and its constitutional effects" (2008) 45 Common Market Law Review 13 at 38.
[137] Dautricourt and Thomas (n 111111) at 434.
[138] A Tryfonidou, "Reverse Discriminination in Purely Internal Situations: An incongruity in a citizen's Europe" (2008) 35 Legal Issues of Economic Integration 43, and updated in "Purely Internal Situations and Reverse Discrimination in a Citizens' Europe: Time to "Reverse" Reverse Discrimination?" at 17, available at http://www.um.edu.mt/europeanstudies/books/CD_MESA09/pdf/atryfonidou.pdf.
[139] UNESCO, citizenship definition, available at *http://www.unesco.org/new/en/social-and-human-sciences/themes/international-migration/glossary/citizenship/*.
[140] *Ibid.*
[141] J Shaw, "The Interpretation of European Union Citizenship" (1998) 61(3) Modern LR 293 at 294.
[142] *Ibid*. at 296.
[143] *Ibid.* at 305.
[144] *Ibid.*

(2) Union citizenship as political participation

Citizenship, it has been argued, is a form of membership relying on rights, duties, and participation in a community.[145] The existence of this status, however, does not suffice to create a European identity on its own,[146] perhaps particularly given that the members of the EU are states, not individuals within those states. Jo Shaw has argued for a citizenship viewed from the perspective of political participation, making the citizen an "active political" one.[147] Some have argued that the very notion of citizenship would lack legitimacy if it were not accompanied by voting rights.[148] This pushes citizenship beyond the merely economically inactive free mover, as is posited in most of the case law. Further, it might be interpreted as requiring some form of commitment from citizens in order to make their status useful, that being participation in local or European elections, the usefulness of rights depending on active participation.

By virtue of the TEU, all Union citizens are conferred the right to vote and stand for election in municipal and European Parliament elections, under the same conditions as nationals of the host member state.[149] The electoral rights gained through European law may be viewed as part of "developing a distinctive European identity",[150] but Shaw goes on to distinguish between participation in local and European elections, arguing that participation in local elections integrates "non-national EU citizens within the host polity" as opposed to creating a European identity.[151] Furthermore, the two differ in their conditions simply because participating in local elections in another member state almost always necessitates movement, whereas voting in European Parliament elections does not. In this regard, the very definitions of "municipal" and "national" were never clarified in the legislation, giving rise to various possible interpretations in light of the differing structures of governance in the member state. This divergence in interpretation appears to undermine any equality between Union citizens,[152] effectively rendering free movement of secondary importance to national political participation.

Participation in local elections, then, does not further any agenda of making people European, in the sense of identity as envisaged by the EU. As such, there is a discrepancy between the objectives of European and local participation, despite their being contained in the same legislative provision, with the right to vote in local elections listed first,[153] a confusing fact given that citizenship, intuitively, would seem closer connected to participation in European elections. Relying on local participation arguably counteracts the feeling of 'Europeanness' by providing closer ties between citizens and their local communities, rather than between citizens and the rest of Europe. If this is so, citizenship matters less as an autonomous status, and rather more as a means of political integration into individual communities within the EU, facilitated by the simplified entry and residence. This does not strengthen the notion of European citizenship, nor does it fit the objective of creating a European identity. Even where EU citizens are granted rights in a host member state, they often do not "share the same identity basis" as most of the national population.[154]

---

[145] Konstadinides (n 6) at 402.
[146] A Schrauwen, "Sink or Swim together? Developments in European Citizenship" (2000) 23(3) Fordham International Law Journal 778 at 793.
[147] J Shaw, *The Transformation of Citizenship in the European Union Electoral Rights and the Restructuring of Political Space* (2007) at 48.
[148] D Kochenov, "Free Movement and Participation in the Parliamentary Elections in the member state of Nationality: An Ignored Link" (2009) 16(2) Maastricht Journal of European and Comparative Law 197 at 198.
[149] Art 22 TEU.
[150] Shaw (n 147) at 48.
[151] *Ibid.* at 49.
[152] Kochenov (n 148) at 209.
[153] Art 22(1) TEU.
[154] M Jakobson and L Kalev, "Transnational Citizenship as Status, Identity and Participation: Comparative Assessment", in Kyriakos N Demetriou (ed), *Democracy in Transition: Political Participation in the European Union* (2013) at 203.

Some have suggested that a European citizenship must imply a "European society".[155] Given that this is demonstrably not the case, the interaction between local, national, and European rights creates a "schizophrenic citizen,"[156] who has many overlapping rights and obligations when it comes to political participation. The EU is a union between the peoples of Europe,[157] and does not as such create a "European society".[158] If EU citizenship is linked to a society or a notion of a *people*, then we cannot sensibly have EU citizenship until, or if, significant social and cultural differences between countries, and societies, are eroded. A *people* seems to imply further social and cultural ties than citizenship, and there are arguably very significant differences between, for example, Swedish and Portuguese nationals, even if both are European citizens. On the other hand, if we view citizenship as being distinct from a society or a people, EU citizenship becomes less puzzling. If, as Shaw argues, citizenship is inherently linked to political participation as opposed to economic participation, then we can sensibly conceive of Union citizenship as political participation, potentially viewed in the widest sense of creating a "connection with the host member state".[159] Even in this case, however, the issue of a link, and therefore generally the crossing of a border or movement, persists, particularly with regards to participation in municipal elections.

(3) Union citizenship as fundamental

The Court's language in *Grzelczyk*, suggesting that Union citizenship was to become fundamental,[160] has been deemed a "rhetorical flourish", but an essentially "empty statement,"[161] lacking the "certainty and clarity" of the early free movement objectives.[162] This in itself is significant, in that it contextualises the generally forceful attitude towards protecting citizenship, and the seemingly sporadic development of the status that the Court has employed. As it stands, the conferral of Union citizenship rests with each individual member state,[163] and is as such not the responsibility of any "central authority."[164] There is something contradictory in having a "fundamental status" attached to a supranational entity governed independently by its parts, undermining the "uniformity of EU law."[165] The somewhat mitigating element of this in *Rottman*,[166] where the CJEU found competence in the area because the national measure would have the effect of actually depriving a Union citizen of that status, [167] does not alter the fact that the actual granting of the status remains with the member state. Additionally, some have argued it is a mistake to prioritise one identity over another in this context, as there is no apparent conflict between national and European identities,[168] which appears to be the implication of the Court pushing the notion of a "fundamental status". The relationship between one individual's citizenships is relevant here, as well as the issue of disenfranchisement referred to above, as one possible disincentive to movement. These issues may amount to prioritisation of one status over the other, whether consciously or not. As such, even where there exists no conflict between national and European citizenship, there may still exist prioritisation.

---

[155] M Everson, "The Legacy of the Market Citizen" in J Shaw and G Moore (eds), *New Legal Dynamics of European Union* (1995) at 76.
[156] *Ibid.* at 77.
[157] Preamble to TEU.
[158] Everson (n 155) at 76.
[159] Shaw (n 147).
[160] Case C-184/99 *Grzelczyk v Centre public d'aide sociale d'Ottignies-Louvain-la-Neuve* [2001] ECR I-06193 at para 31.
[161] Shaw (n 147) at 42.
[162] O'Leary (n 12) at 65.
[163] Everson (n 155) at 74.
[164] Schrauwen (n 146) at 788.
[165] Konstadinides (n 6) at 403.
[166] Case C-135/08 *Rottman v Freistaat Bayern* [2010] ELR I-01449.
[167] *Ibid*. at para 56.
[168] W Sadurski, "European Constitutional Identity?" EUI Working Papers, LAW No. 2006/33 (European University Institute, Department of Law) at 21.

(4) Credibility and aims of the Court

The notion of European citizenship, and the fact that its protection appears bound by more conditions than national citizenship, becomes less credible as a useful legal status where the rights offered by this status have strings of conditions attached to them, the application of which has not always been clear. Ian Ward has argued that the EU has neither an obvious "textual identity" nor a "determinable political determination."[169] This is perhaps not entirely surprising, seeing as the EU is a unique organisation and its exact purpose shifts alongside various political attitudes. Ward goes on to claim that there is a "lack of legal personality or a constitutional identity" within the EU.[170] He quotes Dierdre Curtin, stating that Maastricht's introduction of a 'Union' only managed to "retard the identification of a European Constitution."[171] Whether or not this perceived lack of identity affects citizenship, Union citizenship cannot exist in a vacuum. Concretely, it exists by virtue of primary legislation in the form of the Treaties, the Court's continued deployment of citizenship as protection, and the existence of national citizenship within the individual member state.

(a) Enfranchising citizens

In *Spain v UK*[172] the Court held member states had the competence to determine who was entitled to vote and stand as a candidate in European elections, in line with Community law.[173] Shortly after, the Court decided *Eman and Sevinger*,[174] where it concluded that the citizenship provisions of the Treaty do not confer "an unconditional right to vote and stand as a candidate" in European elections.[175] In that case, however, the Dutch legislation treated differently people living in different overseas countries and territories, and was thus prevented by the principle of equality.[176]

Delvigne[177] concerned prisoner enfranchisement in France, where legislation provided that people convicted of a criminal offence would lose their civic rights.[178] The French, Spanish and British governments considered that the case was outside the Court's jurisdiction, as no connection with EU law had been established.[179] The Court reiterated that the definition of those entitled to vote was left to member states,[180] but continued by stating that such provision must be considered as "implementing EU law" in the meaning of the Charter.[181] It accepted that such a deprivation may be justifiable, provided that it respects "the essence of those freedoms and, subject to the principle of proportionality, are necessary and genuinely meet objectives of general interest."[182] The restriction was considered acceptable because it respected the essence of the right,[183] and proportionate because it took into account the "nature and gravity" of the offence.[184]

In *Spain v UK*, the Court highlighted the wide margin of appreciation afforded to member states in determining voting rights, as long as the provisions do not restrict the right such that it

[169] I Ward, "Identity and Difference: The European Union and Postmodernism" in J Shaw and G Moore (eds), *New Legal Dynamics of European Union* (1995) at 15.
[170] *Ibid.* at 20.
[171] *Ibid.*
[172] Case C-145/04 *Spain v United Kingdom* [2006] ELR I-07917.
[173] *Ibid.* at para 78.
[174] Case C-300/04 *Eman and Sevinger v College van burgermeester en wethouders van Den Haag* [2006] ELR I-08055.
[175] *Ibid*. at para 52.
[176] *Ibid*. at para 61.
[177] Case C-650/13 *Delvigne v Commune de Lesparre Medoc and Prefet de la Gironde,* 6 October 2015.
[178] *Ibid*. at para 5.
[179] *Ibid*. at para 24.
[180] *Ibid*. at para 31.
[181] *Ibid*. at para 33.
[182] *Ibid.* at para 46.
[183] *Ibid*. at para 48.
[184] *Ibid*. at para 49.

impairs "its very essence and deprive[s] [it] of its effectiveness."[185] Van Eijken and Rossem contended, regarding *Delvigne*, that the Court has now framed voting to the European Parliament as a "subjective fundamental right" of all EU citizens.[186] This may constitute a substantial achievement of citizenship in terms of granting extensive substantive rights, particularly to a vulnerable group such as prisoners. It is argued, however, that the decision remains limited in scope, given the Court's continued acceptance of member state limiting prisoner eligibility to vote in certain circumstances, as well as the sensitive nature of areas such as prisoner voting.[187] If the Court was indeed wary of political tensions in its decision in *Dano*, this may indicate a possibility for caution again.

(b) Cautious development

Delvigne suggests a possibility for EU involvement in situations that *prima facie* appear outside its scope, as well as a potential for Union citizenship to protect citizens even without movement, but it is not clear how this would apply in cases of different subject matter. The decision does not represent a departure from the existence of any connection with EU law as regards citizenship. Such a case would remove what is essentially European about said citizenship, and clearly constitute an undue expansion of the powers of the EU. Rather, the Court made clear that because elections to the European Parliament must be by "direct universal suffrage", decisions as to voting eligibility must fall within EU law.[188] It is contended in that vein that *Delvigne* contributes to the notion of citizenship as political participation, as it regards specifically the right to vote in elections to the European Parliament.

It would seem that the Court is open to pushing the scope of citizenship in order to avoid leaving certain vulnerable groups outside its protection. Such a notion would further support the contention that free movement and other rights are mainly for the "privileged".[189] However, it must be highlighted that both cases, and in *Zambrano's* case, the following decisions concern specific situations, whereby the applicants would otherwise have suffered disproportionately. In *Delvigne*, although the restriction was found permissible, if the Court had decided it did not have jurisdiction, there may have been the potential for abuse by allowing any person convicted of a criminal offence to be denied electoral rights. In any event, the European Court of Human Rights (ECtHR) has held previously that a blanket restriction on convicted prisoners to vote, applied automatically, is a violation of the Convention.[190] Completely ignoring the issue may have created a confusing disjunction between ECtHR and CJEU case law. In *Zambrano*, as has been discussed, removal of the parent third-country national would have entailed the loss of residence in the Union of the citizen child, and negatively impacted their ability to exercise their rights and freedoms conferred upon them by virtue of their status as Union citizens. This appears sensible; if the Union cannot protect against the loss of its own protection and rights, there seems little point in extending those rights further without any additional safeguards. Falling short of this, however, it does not seem as though the Court would apply *Zambrano* to any case that was not of such a serious nature, nor would it be likely to encroach upon member states' competence to determine electoral rights within their own territory.

E. CONCLUSION

The Court seems to have become significantly more open in how broadly it construes the protection offered by Union citizenship, both to those who hold the status, and those that derive rights from

[185] Case C-145/04 *Spain v United Kingdom* [2006] ELR I-07917 at para 94.
[186] H Van Eijken and JW van Rossem, "Prisoner disenfranchisement and the right to vote in elections to the European Parliament: universal suffrage key to unlocking political citizenship?" (2016) 12(1) European Constitutional LR 114 at 130.
[187] *Ibid.*
[188] Case C-650/13 *Delvigne v Commune de Lesparre Medoc and Prefet de la Gironde*, 6 October 2015 at para 33.
[189] Dautricourt and Thomas (n 111) at 148.
[190] *Hirst v UK* (No.2) no 74025/01, (ECtHR) 6 October 2005 at para 82.

Union citizens. It would appear, in that regard, that the Court has moved on from the economic focus it displayed in earlier cases, in order to create a wholesale protection to as wide a range of beneficiaries as possible. There is, however, some inconsistency in how the Court has been willing to do so, as evidenced by the restrictive approach taken in *Dano*, and potentially in the sudden re-emergence of *Zambrano*.

The cases in which protection has been granted without an evident cross-border link seem to be particular and in specific circumstances. While these broaden the scope of protection in certain areas, they do not necessarily contradict the notion that a cross-border situation, or element, is necessary. In its creativity and desire to protect, however, the Court has enumerated conditions under which generosity such as that in *Zambrano* might be extended. As the circumstances refer to the loss of any useful effect of Union citizenship, it is submitted that such protection is unlikely to cover anything less than such situations. For this reason, the result of *Zambrano* and similar cases cannot be taken as conclusive as regards extending protection to seemingly internal situations, including those involving so-called static citizens.

The ramifications and credibility of Union citizenship case law undermine the legitimacy of the Court's claim that the status is, or will become, fundamental. This is so, provided that nothing significantly changes in the legislation, particularly the explicit wording of Directive 2004/38. The EU as an institution is not designed to fit a citizenship based on national citizenship, something which seems clear in light of the varying stances of the Court, as well as the many integral aspects of *normal* citizenship that are lacking in this context. The outcome appears to be a new form of citizenship which has not been fully elaborated upon and compartmentalised in relation to all other rights and freedoms of the EU. It is a status which is at best sporadic, but hardly fundamental in any meaningful sense of the word.

CAN THE CONTRASTING STANDARDS FOR STATEHOOD PUT FORTH BY THE DECLARATORY AND CONSTITUTIVE THEORIES OF RECOGNITION BE RECONCILED? AN EXAMINATION OF KOSOVO'S DISPUTED STATEHOOD

Shruti Venkatraman[*]

A. INTRODUCTION

The international responses to Kosovo's unilateral declaration of independence in 2008 revitalised the debate surrounding statehood and recognition, making it a key case for consideration as our contemporary understanding of the subject in international law continues to evolve. To date, Kosovo is recognised by 114 countries including most European Union (EU) and North Atlantic Treaty Organisation (NATO) member states.[1] The two main theories of recognition are the declaratory theory, which holds that recognition is an acknowledgement of pre-existing legal capacity, and the constitutive theory, which holds that recognition is a necessary precondition for international legal personality.[2] These theories clearly put forward contrasting standards, resulting in there being little clarity or consensus on state creation and secession. In analysing the discrepancies between these theories and by applying them specifically to the case of Kosovo, it becomes clear that neither theory fully explains or justifies the international response both in support of and against an independent Kosovo, and that the two schools of thought cannot be reconciled or merged bearing in mind patterns in state practice. Instead, the forces of politics and international relations dictate how states choose to apply these theories, particularly the constitutive theory, in their response to emerging states.

B. THE DECLARATORY THEORY AND KOSOVO

The declaratory theory views recognition as a political act that acknowledges a pre-existing state of affairs and legal capacity, with statehood being conferred by the workings of international law. This is rooted in Article 1 of the Montevideo Convention, which stipulates that "the state as a person of international law should possess the following qualifications: (a) a permanent population; (b) a defined territory; (c) a government and (d) a capacity to enter into relations with other states".[3] Where these criteria are satisfied, the state will be entitled to the pertaining rights under international law, regardless of whether other states recognise them. For example, in the *Tinoco Arbitration*,[4] it was held

[*] LLB Student, University of Edinburgh.
[1] The Republic of Kosovo's Ministry of Foreign Affairs, International Recognitions of the Republic of Kosovo, available at *www.mfa-ks.net/?page=2%2C224.*
[2] M Dixon, *Textbook on International Law,* 7th edn (2013) ch 9.
[3] Article 1 of the Montevideo Convention on the Rights and Duties of States 1933.
[4] *Tinoco Arbitration, Great Britain v Costa Rica* (1923) 1 UN Rep International Arb Awards 369.

that the fact that the United Kingdom did not recognise Costa Rica's Tinoco government did not preclude it from making claims against the government.

According to the declaratory theory, Kosovo must meet all of the Montevideo criteria. Kosovo can be said to have a permanent population, given that the Convention does not stipulate time constraints around the stabilisation of the population after mass migration and displacement.[5] Kosovo has a defined territory, because despite the province's changing constitutional status, it has always had clear borders.[6] Kosovo has the capacity to enter into relations with other states, as evidenced by its diplomatic relations, most notably with the United States and the EU through the Stabilisation and Association Agreement (SAA).[7] However, it is unclear whether it has an effective government over the extent of its territory. Until 2008, other countries maintained a major supervisory role and currently peacekeepers from NATO and the EU remain in Kosovo,[8] along with some forces from the United Nations Interim Administration Mission in Kosovo (UNMIK).[9] The landmark case on independent governance is the *Austro-German Customs Union Case*,[10] in which it was held that independence requires sovereign existence and freedom from foreign control.[11] However, despite this ambiguity, Kosovo has been recognised by the majority of UN member states. The declaratory approach is partially supported by an analysis of the language used by the governments of Canada, Belgium and Italy in response to Kosovo's declaration of independence, which explicitly states that Kosovo's statehood is a reality.[12]

C. THE CONSTITUTIVE THEORY AND KOSOVO

The constitutive theory holds that recognition is a necessary precondition for international legal personality.[13] In the national context, this was discussed in the *International Registration of Trademark (Germany) Case*,[14] in which it was held that trademarks from East Germany would not be protected under law in West Germany, as the government did not recognise East Germany's statehood. This highlights that the constitutive theory emphasises political acts over legal principles, and encourages relativism that makes international consensus on statehood difficult to achieve.[15]

With regard to Kosovo, the majority of states adopted constitutive approaches of recognition. The International Court of Justice (ICJ) Advisory Opinion held that the declaration of independence did not violate international law,[16] thereby not addressing the question of statehood itself and allowing for other states to adopt subjective criteria for recognition. This was followed by statements by countries that recognised Kosovo, in particular Canada, qualifying their endorsements by stating that the situation was *sui generis* and did not create precedent in order to discourage secession movements.[17] This clearly shows the immense effect of political motives, including that of ensuring

---

[5] M Redman, "Should Kosovo Be Entitled to Statehood?" (2002) 73(3) The Political Quarterly 338.
[6] *Ibid.*
[7] US Department of State, "US Relations with Kosovo", available at *https://www.state.gov/secretary/20172018tillerson/remarks/2017/02/267637.htm*.
[8] BBC News, "Kosovo Profile" (2017) available at *www.bbc.co.uk/news/world-europe-18328859*.
[9] United Nations Ten Stories, "Kosovo: The Untold Story of Diplomatic Breakthrough" (2008), available at *www.un.org/en/events/tenstories/08/kosovo.shtml.*
[10] *Customs Regime Between Germany and Austria*, Advisory Opinion (1931) PCIJ (ser A/B) No 41 Sept 5.
[11] *Ibid.*
[12] R Taillefer, "Recognizing Kosovo: Theoretical and Practical Implications for Recognition Theory and the International Community" (2011), available at *arno.uvt.nl/show.cgi?fid=120523*.
[13] Dixon, *International Law* ch 9.
[14] *International Registration of Trademark (Germany) Case* (1959) 28 ILR 82.
[15] Dixon, *International Law* ch 9.
[16] Advisory Opinion on Accordance with International Law of the Unilateral Declaration of Independence in Respect of Kosovo, International Court of Justice 22 July 2010.
[17] CBC News, "Canada Recognizes Kosovo, Serbia Pulls Ambassador" (2008), available at *www.cbc.ca/news/world/canada-recognizes-kosovo-serbia-pulls-ambassador-1.745469*.

stability and peace in the Balkans. Additionally, an analysis of the language used in response statements by the governments of the United States, Switzerland and Germany indicate a theme of willingness of these nations to enter into diplomatic relations, a suggestion that the recognition itself confers legal status.[18] More importantly, the international community's emphasis on seeking wider recognition, such as by Switzerland calling for clarification[19] and the Ministry of Foreign Affairs in Kosovo making lobbying for recognition a priority,[20] show that the constitutive theory aligns most closely with contemporary state practice and general perceptions of recognition.

D. CONCLUSIONS

Clearly, neither theory fully explains the international response to Kosovo's independence. Despite the use of language that suggests the declaratory theory is being applied in state practice, the use of discretion and overlooking the question of whether Kosovo has legal status under the Montevideo Convention show that this model does not apply. When examining the rhetoric and the situation through the lens of the constitutive theory, it becomes clear that the use of subjective criteria tied to geopolitical issues, such as the stability of the region and the unprecedented level of international involvement, played a significant role in how states responded. The fact that legal consequences, in the form of diplomatic relations, arose from recognition is evidence that the constitutive theory is being applied to some extent. Equally, non-recognition, such as that by Russia and Serbia, is treated as a way to withhold the rights that would be available to a recognised state.[21] However, the main flaw in the application of constitutive theory is that is does not lead to uniformity, because this would mean that Kosovo is viewed as a state by some states and not others. Ultimately, this requires a distinction to be drawn between *de jure* and *de facto* recognition.[22] In this case, it appears that Russia and Serbia acknowledge the *de facto* existence of Kosovo and not its *de jure* existence, because it urges the United Nations to take severe administrative action against Kosovo's government.[23] As in the *Tinoco Arbitration,*[24] this is also an application of the declaratory theory.

The case of Kosovo's disputed statehood and the continuing lack of clarity on its legal status highlights the deficiencies of theoretical approaches to recognition, and the difficulty of reconciling the two schools of thought. The declaratory theory provides too simplistic a view while the constitutive theory provides a flexible approach but lacks cohesiveness. The constitutive theory aligns more clearly with state practice, largely due to the influence of political motivations which encourage states to use discretion and apply subjective criteria in line with key national and international considerations. There have been attempts to link the two theories, most notably by Hersch Lauterpacht QC, who maintained that states have a duty to grant recognition to any new state that meets the conditions of international law.[25] However, Kosovo's case shows that such an intermediary theory does not fully apply because the distinction between *de jure* and *de facto* recognition is not sufficiently addressed and because state practice is largely guided by political rather than legal principles. This shows that the two theories cannot be reconciled, and that where both theories are applied, the approach is less formalised and straightforward than the Lauterpacht doctrine. This lack of clarity in the legal order acts as a fundamental impediment for reaching a consensus in the international community on this matter. To an extent, this ambiguity also arises from the larger struggle that undermines the effectiveness of international law, which Raic describes as "the absence

[18] Taillefer (n 12).
[19] Statement by the President of the Swiss Confederation, Pascal Couchepin, The Swiss Federal Council 27 February 2008.
[20] The Republic of Kosovo's Ministry of Foreign Affairs (n 1).
[21] Taillefer (n 12).
[22] *Ibid.*
[23] BBC News, "Serbia Urges UN Action on Kosovo" (2008), available at *news.bbc.co.uk/1/hi/world/europe/7244333.stm*.
[24] *Tinoco Arbitration, Great Britain v Costa Rica* (1923) 1 UN Rep International Arb Awards 369.
[25] M N Shaw, *International Law*, 8th edn (2017) ch 8.

of a central organ with general powers of attribution and enforcement rights and obligation".[26] Ultimately, as shown through a close examination of the situation in Kosovo, political motivations inform states' responses to emerging states and in the absence of clear universal standards, the two theories of statehood cannot be reconciled to explain or govern state practice in this area of international law.

.

[26] D Raic, *Statehood and the Law of Self-Determination* (2002) 50.

SHOULD STEP-CHILDREN HAVE AN AUTOMATIC RIGHT IN INTESTACY AND IF SO HOW SHOULD THEIR CLAIMS BE REGULATED?

Eilidh MacQuarrie[*]

A. INTRODUCTION

It has been said that succession law straddles the "core institutions of property and family".[1] Therefore, as the arrangement of the average family changes over time, so should the law governing intestate succession. The new Succession (Scotland) Act 2016 introduced changes to the law of succession for the first time since 1964.[2] This Act brought changes, including the rectification of a will in limited circumstances, and, most notably, invalidating any bequest in favour of an ex-spouse or ex-civil partner. Although these changes echoed some of the recommendations addressed in the Scottish Law Commission's (SLC) most recent Report on Succession,[3] the issue of step-children has yet to be tackled.

This article will address whether step-children should be included in future reform proposals for intestacy law. The first section will discuss the policy objectives of our system of inheritance, as well as the effect that a changing society should have on our intestacy rules. Feedback provided by Public Opinion Survey (the "2005 survey") will be analysed and contrasted with the suggestions of the SLC.[4] I will conclude by reviewing the current law surrounding step-children.

The second section will analyse the proposals of the SLC chronologically, focusing on their 2009 Report which did not recommended the introduction of intestacy rights for step-children. The

[*] DipLP student at the University of Edinburgh.
[1] D Reid, "From the Cradle to the Grave: Politics, Families and Inheritance Law" (2008) 12 (3) Edin LR 391 at 392.
[2] Hereafter S(S)A 2016.
[3] Report on Succession (Scot Law Com No 215, 2009).
[4] Scottish Executive, "Attitudes Towards Succession Law: Findings of a Scottish Omnibus Survey" (2005).

third section will look at how a system of step-children's intestacy rights would function. It will specifically consider whether the law should have a fixed or discretionary definition of a step-child and whether intestacy rights should be fixed or discretionary.

B. SETTING THE SCENE

(1) Social policy and intestate succession

(a) Guiding principles of intestate succession

In 2008, Reid stated that the way in which "a society restricts or encourages freedom of ownership and the accumulation of wealth reflects the values of that society".[5] It is undeniable that a society's values should influence succession, but the lack of legislation in the area until 2016 infers that the Scottish Parliament is reluctant to make any drastic changes. It is possible that this is because the area is politically sensitive. Intestate succession law affects most Scottish citizens over the course of their lifetime, making it more personal than other areas of the law. Any change made to intestacy law will impact every Scottish person without a will. It is therefore important that the law is appropriate for the majority of Scotland. On this basis, a substantial change is more likely to be met with criticism from academics, lawyers and the public where it does not appear to suit the majority of Scotland.

Academics have commented on the fact that Scots succession law follows a set of guiding principles.[6] These principles are not codified and can only be identified through their latent presence in legislation and academic commentary. Burns touched on the guiding principles or "policy objectives" of Scots and English succession law in 2013.[7] She identified "efficiency, simplicity or fairness between family members" as factors to be considered when looking at law reform.[8]

(b) Efficiency

Administrative efficiency refers largely to minimising litigation, as it has the potential to waste a considerable portion of the deceased's estate on legal fees. The law tries to avoid uncertainty in legislative provisions as it can lead to litigation determining the meaning of a provision in court.

One example of Scots law avoiding uncertainty would be cohabitants' rights. Scotland currently recognises cohabitants' rights under Section 9 of the Family Law (Scotland) Act 2016;[9] however, these rights are discretionary and require an application to be made to the court within six months of the death of the cohabitant.[10] The present system of cohabitants' rights has been criticised by Guthrie and the SLC.[11] Guthrie identified lack of legal certainty and clarity as the "fundamental difficulty with the legislation".[12] The SLC echoed this analysis and was critical of the lack of guidance given to the court in addition to the "unfettered extent of the court's discretion".[13] The criticism of the current system of cohabitants' rights suggests that administrative efficiency – specifically certainty – is fundamental when considering the reform of Scots succession law.

---

[5] Reid (n 1) at 391.
[6] F Burns, "Surviving Spouses, Surviving Children and the Reform of Total Intestacy Law in England and Scotland: Past, Present and Future" (2013) 33 (1) LS 85.
[7] *Ibid.* at 97.
[8] *Ibid.*
[9] Hereafter, FL(S)A 2006.
[10] Family Law (Scotland) Act 2006, s 29(6); cf *X v A (No 1)* 2016 SLT (Sh Ct) 404.
[11] T Guthrie and H Hiram, "Property and Cohabitation: Understanding the Family Law (Scotland) Act 2006" (2007) 11(2) Edin LR 208; Report on Succession (Scot Law Com No 215, 2009) para 4.7.
[12] Guthrie & Hiram (n 11) at 213.
[13] Report on Succession (Scot Law Com No 215, 2009) para 4.7.

(c) Simplicity

Simplicity was the second guiding principle identified by Burns. If the law is simple, it has the benefit of being read and understood by lay people. Simplicity is identified as fundamental by academics in other areas of the law, such as taxation,[14] where is lowers administrative costs.

A home-made will is a testament created by a testator without the assistance of a solicitor, often using a will-making kit. A home-made will is valid where it satisfies the requirements set out in the Requirements of Writing (Scotland) Act 1995.[15] Therefore, for succession purposes, simplicity is desirable to ensure that a lay person understands the legislation and knows whether they would like their estate to be distributed in accordance with the intestacy principles. This means that if they are unhappy with how intestacy law would affect them, they can make a will. It is, therefore, important that succession legislation is as simple as possible, thus allowing people to understand their position without the need of a solicitor.

(d) Fairness of distribution

The final guiding principle identified by Burns was fair distribution and SLC recognised this as the primary purpose of reform.[16] It concluded that simplicity is a tool that should be used to achieve this,[17] for example, through the removal of the distinction between heritable and moveable property. In practice, however, it is a broad principle that is more difficult to define and achieve than simplicity or efficiency. It is a more sophisticated concept that does not have a singular definition, but rather is a subjective issue. To achieve fair distribution in a modern context, Scots succession law should reflect societal changes and ought to be considered fair by the majority of those affected by the law.

It is possible to have an efficient system that is unfair, and it is also recognised that the fairest system is unlikely to be simple. When considering future reform, it is important to balance these principles in a way that results in a system that is the best for society. Ultimately, the guiding principles of intestacy are useful tools for reform, but they are merely advisory.

(e) The effect of a changing society

Outside of Burns' guiding principles, the SLC, in its 2007 Discussion Paper,[18] recognised that society and the "principles to which each jurisdiction gives different weights" shapes the law of intestate succession.[19] Since the Succession (Scotland) Act 1964 (S(S)A 1964) was brought into force civil partnerships have been introduced, gay marriage has been legalised, cohabitants have gained limited recognition as discretionary beneficiaries, reconstituted families have increased in prevalence and the number of divorces in Scotland has increased from 2455 to 9030 in 2014-15.[20] Norrie noted that the SLC's Discussion Paper was timely due to the changing understanding of a family.[21]

---

[14] Institute for Fiscal Studies, "The Mirrlees Review, Tax by Design" (2011) at 42, available at *https://www.ifs.org.uk/docs/taxbydesign.pdf*.
[15] Requirements of Writing (Scotland) Act 1995, s 1(2)(c).
[16] Report on Succession (Scot Law Com No 215, 2009) para 2.3.
[17] Report on Succession (Scot Law Com No 215, 2009) para 2.3.
[18] Discussion Paper on Succession (Scot Law Com DP No 136, 2007).
[19] *Ibid.* para 2.2.
[20] National Records of Scotland, "Divorces Time Series Data" (2013), available at *www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/vital-events/vital-events-divorces-and-dissolutions/divorces-time-series-data;*
The Scottish Government, "Civil Justice Statistics in Scotland 2014-2015" (2016), available at *www.gov.scot/Resource/0049/00497242.pdf*.
[21] K McK Norrie, "Reforming succession law: intestate succession" (2008) 12(1) Edin LR 77.

The law of intestacy in Scotland has traditionally involved blood relationships or their legal equivalents such as adoption and marriage.[22] It should be remembered, however, that until recently, families themselves were defined by blood relationships because divorces were less frequent. The introduction of cohabitants' rights in the FL(S)A 2006 has demonstrated the possibility of extending intestacy rights beyond the nuclear family model which could result in a system that better reflects modern Scottish society.[23] Thus, the next step, it is suggested, is the introduction of intestacy rights for step-children. Before the Scottish Law Commission's recommendations can be analysed, it is important to look at the views of the population through the Public Opinion Survey.[24]

(f) Public understandings of intestate succession

Intestacy provisions only apply to Scottish citizens who die without having made a valid testament. The most recent statistics on will making in Scotland were collected by the Scottish Consumer Council over a decade ago (the "O'Neill Report").[25] It found that only 37% of the population had created a will,[26] but there was a noticeable variation when looking separately at age and socio-economic background. Unsurprisingly, 69% of the sample group aged over sixty five had a will, which is a much higher proportion than the overall data set.[27] This suggests that as people age, they tend to think more seriously about what happens after they die.

46% of the respondents without a will claimed that they had not made testamentary provision because they had "never got around to it".[28] Another 28% claimed that they had not even thought about it.[29] Of the respondents that had made a will, 17% claimed they had done so because they had a child or children.[30]

The O'Neill Report also tested the respondent's knowledge of intestacy law provisions. One of the scenarios involved the rights of step-children. Only 50% of the respondents correctly identified the law in this area and 37% believed that children have equal rights to step-children upon intestacy.[31] Although this cannot be stated conclusively, there is a suggestion that some step-parents may not write a will as they incorrectly believe the default laws of intestacy will successfully fulfil their wishes. In other cases, they may intend to make a will but are unaware of the urgency of doing so because they falsely believe that the intestacy provisions would be an appropriate fall-back. It may be unsurprising that so many respondents believed this as the Scottish Executive 2005 Survey revealed that 75% of its respondents were in favour of step-children having at least some rights to their step-parent's estate.[32] Furthermore, 68% of respondents believed that step-children should be treated as equally as their step-parent's natural children.[33]

It could be argued that these statistics only represent a cross-section of the population; however, it is still surprising that the figures did not impact upon the SLC's decision to introduce rights for step-children in intestacy. The 2007 Discussion Paper did consider the results of the Scottish Executive 2005 Survey, but concluded that the disadvantages of such a provision outweighed the

[22] Report on Succession (Scot Law Com No 215, 2009) para 2.23.
[23] Family Law (Scotland) Act 2006, s 29.
[24] Scottish Executive, (n 4).
[25] O'Neill, "Wills and Awareness of Inheritance Rights in Scotland" (2006), available at http://webarchive.nationalarchives.gov.uk/20090724140641/http://scotcons.demonweb.co.uk/publications/reports/reports06/rp10wrep.pdf.
[26] *Ibid.* at 6.
[27] *Ibid.*
[28] O'Neill (n 25) at 8.
[29] *Ibid.*
[30] O'Neill (n 25) at 9.
[31] O'Neill (n 25) at 18.
[32] Scottish Executive (n 4) at 7.
[33] Scottish Executive (n 4) at 8.

advantages.[34] This stance was echoed by the subsequent Report.[35] The SLC's specific arguments on this point will be discussed in Section C.

(2) Current legal framework

At present, for a step-child to be an heir on the death of their step-parent, specific testamentary provision must have been made. The obvious solution may be to encourage step-parents to make wills but this is unrealistic considering that around 60% of the population have not made wills.[36] An alternative solution may be to encourage step-parents to adopt their step-children. However, this may not always be desirable as it would result in a step-child losing succession rights to the estate of their natural parent as they would then only be treated as the biological child of their adoptive parent.[37]

(a) Recognition of step-children in Scots law

Last year, the S(S)A 2016 brought law reform to succession law for the first time in over fifty years. However, no changes were made to the rights of step-children despite support for intestacy rights for step-children being voiced in the Scottish Executive 2005 Survey.[38] Although step-children are not currently recognised in succession law, they are recognised in other areas of Scots law. The Family Law (Scotland) Act 1985 recognises an obligation of aliment owed by a person to a child accepted as a child of their family.[39] This definition is broad enough to cover step-children, whereas the definition found in the S(S)A 1964 is restricted to natural children.[40]

Step-children are also recognised under the Damages (Scotland) Act 2011. A "relative" is defined in the Act as:

> a person who is […] a parent or child of the deceased, accepted by the deceased as a child of the person's family or was accepted by the deceased as a child of the deceased's family.[41]

This allows step-children to claim damages from the wrongdoer whose actions caused the death of a step-parent.

As step-families are commonplace in our society, it begs the question why step-children are not yet provided for in Scots intestacy law. This is highlighted by the great value placed on blood relationships in succession law. This statement can be shown by the fact that whole blood siblings take precedence over half-blood siblings under Section 3 of the S(S)A 1964. However, families are becoming less reliant on blood relationships and are focusing more on status. This can be shown by the increased prevalence of gay marriage, adoption, foster-parenting and step-parenting. It would therefore be a logical step to extend the law of intestacy to step-children.

Following the guiding principles of succession law, certainty is valued highly within our legal system. Furthermore, consistency across the legal system can lead to increased certainty. Yet, the treatment of step-children as immediate family members across our law varies drastically from discretionary rights in personal injury law to a complete absence of rights in intestacy law. The O'Neill Report has shown that the public's understanding of intestacy law is limited.[42] This may be a result of the inconsistent treatment of step-children across Scots law. Introducing additional rights for step-children would

[34]Discussion Paper on Succession (Scot Law Com DP No 136, 2007) para 2.80.
[35] Report on Succession (Scot Law Com No 215, 2009).
[36] O'Neill (n 2525) at 6.
[37] Adoption Act 1976 s 39.
[38] Scottish Executive (n 4) at 7.
[39] Family Law (Scotland) Act 1985 s 1(1)(d).
[40] Succession (Scotland) Act 1964 s 2(1)(a).
[41] Damages (Scotland) Act 2011 s 14(1)(b).
[42] O'Neill (n 25).

ensure consistency across the legal system and guarantee certainty for step-children. Consistency would be achieved by granting step-children similar rights within succession and family law, where currently the positions differ.

(b) Fairness of distribution and the exclusion of step-children

Since the late nineteenth century, Scots law has prioritised spousal entitlement over the legal rights of the deceased's issue.[43] This is justified on the basis that the spouse tends to be reliant on the deceased as they shared the matrimonial home.[44] It is also becoming increasingly normal for issue to have reached adulthood when their parents die, at which point they tend to be financially secure. There is a further presumption that where the deceased leaves young children, the surviving spouse will look after them and that they are in a better position to do so where they inherit most of the estate. However, this latter assumption may not always apply in circumstances involving step-children.

As spouses are prioritised, there may be a scenario where a spouse's prior rights exhaust the deceased's estate meaning that a natural child is unable to claim their legal rights. Where the spouse is not the parent of the step-child, the step-child is unable to recover any family property from their step-parent if the latter has not made a valid will in their favour. The property inherited by the step-parent would be passed upon intestacy to their own issue and family members in accordance with the Section 2 provisions.[45] Although the Scottish legal system favours the spouse over issue, it still makes provision for issue in the form of legal rights, thus implying that keeping property within the family remains important in the field of intestacy law.

The challenge involving a second spouse is addressed in other jurisdictions, such as the Netherlands which uses a Wilsrecht. The Wilsrecht provides that the entire estate is inherited by the deceased's spouse and any claim by the deceased's children – either natural or step-children – is held in liferent by the spouse.[46] This claim crystallises upon the death or bankruptcy of the spouse.[47] In addition, the position of the step-child is strengthened as they are able to enforce their rights at any time.[48] This means that the spouse is unable to exhaust the share of a step-child during their lifetime. The main benefit of this approach is that a step-child will not lose their claim to their parent's estate when their step-parent inherits everything.

One argument against replicating the Dutch system is the infrequency of a second spouse erasing a child's claim on their parent's estate. Furthermore, as the intestacy rules are default rules, there is no need to legislate against circumstances that are unlikely to affect the majority of the population. The SLC also believed this system relies heavily on liferent, something rarely used in Scotland today.[49] It would therefore be irrational to introduce a solution that would be operated infrequently. It could be submitted that an appropriate solution to protecting a step-child from a second spouse's claim would be to introduce general rights in intestacy for all step-children. This would be more widely applicable and would also protect any step-children in this specific scenario, one which is particularly disadvantageous under Scots law.

(c) Unlimited heirs and step-children

Where a person dies intestate and there is no identifiable family, their estate will pass to the Crown as the ultimate heir – or *ultimus haeres*. Although there has been debate over whether the Crown can be

[43] Burns (n 6) at 85.
[44] Reid (n 1) at 396.
[45] Succession (Scotland) Act 1964 s 2.
[46] Dutch Civil Code Article 4.13.2.
[47] Dutch Civil Code Article 4.13.3.
[48] Dutch Civil Code Article 4.21.
[49] Discussion Paper on Succession (Scot Law Com DP No 136, 2007) para 2.45.

an heir in the "proper sense",[50] it has a right to property where there is no other heir available. In Scotland, this is unlikely to happen due to the system of unlimited heirs. Under the S(S)A 1964 the search for heirs in intestacy extends to ancestors of the deceased beyond grandparents, followed by the siblings of those ancestors before the ancestors of a more remote generation.[51] The result is that there is no restriction on the degree to which a beneficiary is related to the deceased. This type of beneficiary is sometimes referred to as the "laughing heir" in American terminology due to the distance of relationship between the deceased and the heir which only results in minor bereavement for the remote ancestor.[52]

In our current system, a "laughing heir" may inherit instead of a cohabitant or a step-child because the former is deemed to be "legally deserving", whereas the latter is only "morally deserving".[53] There are practical issues with the application of this provision, for example, the cost of tracing the rightful heir. In other jurisdictions, this right is limited by the remoteness of the relationship to the deceased to avoid these difficulties.

The issue of restricting the remoteness of eligible relatives in Scotland was raised in Memorandum 69.[54] It was rejected on the basis that a remote relative is a preferable heir to the State despite the potential lack of a morally deserving relationship. It could therefore be argued that where remote relatives are granted a potential claim on intestacy, succession rights on intestacy should also be introduced for step-children as they are more likely to have had a meaningful relationship with the deceased than a distant relative.

C. CRITICAL ANALYSIS OF THE SLC'S PROPOSALS

(1) Scottish Law Commission reports

The SLC has published three Memoranda, a Discussion Paper and two Reports highlighting many problems with the S(S)A 1964 along with recommendations of how future legislation should attempt to address some of these problems. Within these publications, recommendations have been made to change the legal position of a step-child within intestacy.

This section will discuss the recommendations and overall themes of these reports chronologically before analysing the specific arguments raised that rebut the introduction of step-children's intestacy rights. Specific attention will be paid to the arguments involving adoption and will-making as an alternative solution to intestacy and to the triple parentage argument.

(a) Memorandum 69

The purpose of the Memorandum was to obtain comments on the potential reform of intestate succession law.[55] Although not a final report, it is useful in showing the progression of the Scottish Law Commission's recommendations after the introduction of the S(S)A 1964.

The Memorandum stated that an external suggestion had been made to look at the position of step-children in succession law.[56] However, it was stated that it would not be appropriate to equate the rights of step-children with natural children as a step-relationship differs from the natural relationship

[50] *Goldie v Murray* (1753) Mor 3183 per Lord Kames.
[51] Succession (Scotland) Act 1964 s 2(1)(i).
[52] D Cavers, "Change in the American Family and the 'Laughing Heir'" (1935) 20 Iowa L Rev 203.
[53] K Anderson, "Intestacy in Scotland: The Laughing Heir" (2011) 2(1) Aberdeen Student L Rev 52.
[54] Intestate Succession and Legal Rights (Scots Law Com Consultative Memorandum No 69, 1986).
[55] Intestate Succession and Legal Rights (Scots Law Com Consultative Memorandum No 69, 1986) para 1.1.
[56] Intestate Succession and Legal Rights (Scots Law Com Consultative Memorandum No 69, 1986) para 4.89.

as it can be terminated by divorce.[57] There was, however, support for granting step-children a discretionary right which would be determined by the length of the step-relationship.

The Memorandum also raised the issue of whether a former step-child could apply for a discretionary provision and concluded that they could. It was acknowledged that the factual relationship may not have been terminated between the step-parent and step-child at the point of divorce and that custody may have been awarded to the step-parent. Ultimately, the Memorandum indicated that any discretionary reward should be based on the relationship and should not apply to adult step-children where there was no meaningful step-relationship.

Accepted children were distinguished from step-children in the Memorandum.[58] This category was broader than the step-child classification and included children being looked after by a relative and also the child of a cohabitant. The SLC considered that generosity should not incur the moral responsibility of providing for an accepted child in a will or otherwise.[59] However, it was persuaded to think that any accepted child – whether step-child or otherwise – should be entitled to a discretionary provision.

(b) Report on Succession 1990

The SLC did not directly address step-children's intestacy rights in its 1990 Report but briefly acknowledged that they can inherit from a surviving spouse through the creation of a will.[60] Currently, children risk disinheritance from their natural parent where there is a second spouse and modest estate. This was largely ignored in the 1990 Report. It is possible that this is a result of the political climate at the time because the SLC also failed to recognise controversial relationships, such as cohabitation. This oversight is surprising because step-children and accepted children's rights were included in the Memorandum in 1986 which contributed to the formation of the 1990 Report.

(c) Discussion Paper 2007

The Scottish Executive 2005 Survey was used as a point of reference in the SLC Discussion Paper for the changing perception of step-children in society and attention was turned to whether the law should be changed on this basis. It was recognised that the current provisions can produce "harsh anomalies",[61] and that step-children have been acknowledged in other areas of law.[62]

Intestacy rights were recognised as revolving around "blood relationships" or their "legal equivalents" in a historical context.[63] It was submitted that China was the only jurisdiction, under the scrutiny of the Commission, who conferred such rights on step-children. Secondly, it defended the current legislation on the basis that there are ways to provide for step-children, for example through provision in wills or rights gained through formal adoption. It was also acknowledged that the purpose of including non-biological children in other areas of legislation was the need to protect a step-child where there had been an established close relationship with the step-parent.

The Discussion Paper then considered the practical implications of determining whether a child had been sufficiently "accepted".[64] It was thought that the acceptance of a child does not result in the removal of parental rights and responsibilities of a non-resident parent. The "triple parent"

---

[57] *Ibid.*
[58] Intestate Succession and Legal Rights (Scots Law Com Consultative Memorandum No 69, 1986) para 4.90.
[59] *Ibid.*
[60] Intestate Succession and Legal Rights (Scots Law Com Consultative Memorandum No 69, 1986) para 2.16.
[61] Discussion Paper on Succession (Scot Law Com DP No 136, 2007) para 2.73.
[62] Damages (Scotland) Act 1976; Matrimonial Homes (Family Protection)(Scotland) Act 1981; Family Law (Scotland) Act 1985.
[63] Discussion Paper on Succession (Scot Law Com DP No 136, 2007) para 2.75.
[64] *Ibid.*

argument was introduced which recognised that an accepted child could end up being treated as having three parents and could put them in an advantageous position over a natural or adopted child.

The SLC also highlighted the difficulties of deciding whether an accepting adult should be entitled to succeed to their accepted child's estate and stated that equating natural children to accepted children would be a step too far. An alternative solution was then introduced, whereby an accepted child could apply to the court for an award out of the deceased's estate under the protection from a disinheritance scheme.

The Discussion Paper concluded with several questions pertaining to the introduction of accepted children's rights on intestacy. This was done to elicit views in preparation for the Report on Succession published in 2009.

(d) Report on Succession 2009

The SLC began its report by accepting that there is considerable public support for giving intestacy rights to step-children and echoed the Discussion Paper in admitting that there can be "harsh anomalies".[65] The Report then continued to reiterate the position of the Discussion Paper claiming that Scots succession law relies on blood relationships to ensure certainty. This provided a foundation for their overall rejection of the idea of step-children's intestacy rights.

Firstly, it was stated that introducing non-biological children's intestacy rights would reduce the certainty of the law. This was justified on the basis that it is difficult to establish if the deceased had accepted the child. Furthermore, the Report questioned whether an accepted child would be granted intestacy rights to their step-parent's extended family.

Secondly, an argument was raised that suggested such rights would disproportionately advantage non-biological children who could inherit from two natural parents and a step-parent. This echoed the commentary of the Discussion Paper. It concluded that the disadvantages of a change to the law would outweigh any advantages and highlighted that the simplest solution would be the creation of a will by the step-parent.

(2) Common themes and arguments

Since 1986, the SLC has consistently stated that natural children and non-biological children should not have equal rights. A discretionary system was supported by the Memorandum but was rejected in the later Reports because of its failure to adhere to the simplicity and certainty principles that underpin our current system. The Memorandum was more open to a provision extending intestacy rights to step-children, but this approach has since lost support in publications that have followed.

SLC's primary concern with possible reform is that non-biological children would be advantaged by having a third parent. Its secondary issue with such rights is that there are already mechanisms that allow step-children to inherit from their step-parent, most notably through adoption and valid testation. Finally, it is submitted that succession law has relied on blood relationships in the past because of issues of certainty. These arguments will now be addressed.

D. ALTERNATIVES TO LEGISLATING

The most common argument against reform to intestacy law – both generally and in relation to step-children specifically – is that if a person is unsatisfied with the intestacy provisions, they can create a

[65] Report on Succession (Scot Law Com No 215, 2009) para 2.31.

will. This thread of argument was employed by the SLC in 2009.[66] It is undeniable that this view is logical, but it overlooks the will-making statistics in Scotland.

Another frequently used argument against intestacy reform for step-children specifically is that any step-parent who truly wishes their step-children to be entitled to a claim on their estate could adopt them. Kenneth Reid,[67] and the SLC in their Discussion Paper,[68] have used this argument, but it is not watertight and will be critically reviewed in this section.

(1) Testation as an alternative solution

The standard rebuttal to any intestacy reform is that where a testator is unhappy with the general intestacy provisions they can be overridden with a will. Scotland has protection from disinheritance in the form of legal rights and prior rights, but, aside from these, there is relative freedom of testation. This means that step-parents could include their step-children in a will where their estate has not been exhausted by legal and prior rights. Although it would be an obvious solution to simply encourage step-parents to make wills, there are some fatal flaws with this argument.

The most obvious flaw is that only 37% of people in Scotland have made a will.[69] If this line of argument is to be followed, it would be presumed that the 63% of the population without a will were happy with the current intestacy provisions. However, the Public Opinion Survey showed otherwise, with 75% of respondents expressing favour for some intestacy rights for step-children.[70] Additionally, it was previously highlighted that much of the population is unaware of the intestacy rights of step-children in Scotland.[71] It cannot, therefore, be assumed that when a step-parent is unhappy with intestacy provisions they will create a will. This is largely because step-parents are unaware of the intestacy provisions and do not realise that they require a will to fulfil their wishes.

This stance would be more compelling if the SLC had suggested testation was a solution and that the current intestacy provisions should be advertised through a governmental campaign. With this approach, step-children would be better protected as step-parents would be more aware of the fact that they need to create a will. Although the SLC is not responsible for the creation of policy-based campaigns, simply stating that making a will is a suitable option is a weak justification for the complete rejection of step-children's rights. This is not to say that the creation of wills by step-parents would not solve the issue, but it could be argued that the improbability of this solution warrants legislative intervention.

There is currently a scheme operating in England and Wales where law firms draw up wills for free in exchange for a donation to charity.[72] This has encouraged many testators to create wills – which are more likely to be valid as they are not home-made – and has the effect of ensuring that a testator's intentions are met. If the alternative solution to the protection of step-children in intestacy is will-making, then a similar scheme or campaign could be started in Scotland as a way of helping to address the issue. The main consideration here is whether legislative intervention in this area would be disproportionate. If this is the case, a campaign of this type would be an appropriate alternative, but the academic assertion that will-making is an option does not go far enough to tackle the issues faced by step-children in the first chapter of this discussion.

---

[66] Report on Succession (Scot Law Com No 124, 1990) para 2.33.
[67] K G C Reid, "Intestate Succession in Scotland" in K G C Reid, M de Waal and R Zimmerman (eds), *Comparative Succession Law: Volume II: Intestate Succession* (2015) 371 at 397.
[68] Discussion Paper on Succession (Scot Law Com DP No 136, 2007) para 2.75.
[69] O'Neill (n 25) at 6.
[70] Scottish Executive (n 4) at 7.
[71] O'Neill (n 25) at 7.
[72] Free Wills Month, "Frequently Asked Questions" available at https://freewillsmonth.org.uk/.

(2) Adoption as an alternative solution

Adoption is another solution offered as an alternative to introducing step-children's intestacy rights. It is argued that if a step-parent wants to grant their step-child intestacy rights equal to that of a natural child, they should pursue the legal equivalent of a parent-child relationship under the Adoption and Children (Scotland) Act 2007.[73] However, academics do not appear to have given due attention to the practicalities of adoption when making this assertion. A non-family lawyer or lay person may assume from this solution that obtaining an adoption order for a step-child is a simple means to an end. Unfortunately, this is not the case as there are several hurdles to overcome. Namely, a step-child would have to give up their rights to a birth parent's estate to be able to claim. Adoption should also be considered as a last resort because the legislation does not intend for it to be used for the creation of succession rights.

(*a) Refusal by the step-child*

The most obvious roadblock to adoption as a solution to step-children's succession rights is that any child over the age of 12 can refuse to give their consent to the adoption where the court is satisfied that the child is capable of giving consent.[74] It could be argued that a step-child who wants intestacy rights over their step-parent would not refuse an adoption order. However, family circumstances could be complicated and the reasons for a refusal could be related to the step-child's close relationship with their birth parent. Furthermore, adoption is only an option where a person is under the age of 18 meaning that adoption is only a practical solution where a child is under 12 years old and unable to consent or between the ages of 12 and 18 years old and consenting.[75]

(b) Refusal by a birth parent

A second barrier to adoption is parental consent.[76] Admittedly, if parental consent is not forthcoming, there are several situations when a parent's lack of consent is unnecessary,[77] but they are high barriers to overcome. Where both natural parents are alive,[78] and capable of giving consent,[79] one of the applicable grounds for a natural parent's consent to be dispensed of is the welfare of the child.[80] Alternatively, consent can be overridden where a parent with parental responsibilities and rights is unable to satisfactorily exercise those rights and will "continue to be unable to do so".[81] In scenarios involving a step-parent granting their step-children intestacy rights, it is possible that both natural parents are still alive and exercising their parental rights and responsibilities. Without the natural parent's consent, an adoption order is difficult to obtain.

In *X v Y*,[82] a step-mother (X) applied for an adoption order in respect of her step-son who lived with her and the boy's father. There had been contact between the step-son and his birth mother until 2011, at which point he had expressed a desire to cease contact. Even though the step-son and his birth father consented to the adoption order, the order was refused. There had been no evidence suggesting the birth mother was unable to exercise her parental rights and responsibilities. This case illustrates that if a step-child has two natural parents that are able to appropriately exercise their parental rights and responsibilities and do not consent to the adoption, a step-parent's claim will be rejected. Ultimately, adoption is not appropriate as an alternative solution to the introduction of step-children's intestacy rights.

---

[73] Hereafter, the AC(S)A 2007.
[74] Adoption and Children (Scotland) Act 2007 s 32(1).
[75] Adoption and Children (Scotland) Act 2007 s 28(4).
[76] Adoption and Children (Scotland) Act 2007 s 31(2)(a).
[77] Adoption and Children (Scotland)Act 2007 s 31(2)(b).
[78] Adoption and Children (Scotland) Act 2007 s 31(3)(a).
[79] Adoption and Children (Scotland) Act 2007 s 31(3)(b).
[80] Adoption and Children (Scotland)Act 2007 s 31(3)(d).
[81] Adoption and Children (Scotland) Act 2007 s 31(4)(c).
[82] *X v Y* 2015 Fam LR 41.

(3) Triple parentage

The triple-parent argument suggests that by introducing intestacy rights for step-children, the law would allow step-children to claim from the estates of three parents. This argument is supported by the fairness of distribution principle on the basis that it would be undesirable to put step-children in an advantageous position over natural and adopted children.

This argument has academic support from the SLC and a number of jurists. Reid claimed the only alternative to triple parentage would be replacing a claim against a natural parent with a claim against a step-parent resulting in a step-child being limited to inheriting from two parents.[83] However, he deemed this unacceptable and stated that the current law was fit for purpose as step-parents were still able to adopt their step-children or include them in a testament. Norrie also acknowledged the triple-parent argument in 2008, though he did not confine the problem to three parents, but stated that a larger number of inheritances was possible.[84] Norrie claimed that the more reconstituted a family was, the more unequal the claims between step-children would become and that this would be a "bad social policy".[85]

(a) The guiding principles

When considering the guiding principles, a system allowing a child to have three parents could be simple and efficient but may not fall in line with the principle of fair distribution. As spouses are prioritised in intestate succession, their rights would not be affected by the introduction of a step-child's claim.

Therefore, the claimant most likely to be affected by such rights, if any, is the deceased's issue as it is possible that a step-child's intestacy right could reduce the value of a biological child's claim. Although this is a strong justification for abandoning the pursuit for step-children's intestacy rights, where step-children and biological children are raised as equals within a family unit, the biological child may agree that the step-child should be entitled to a claim on the estate.

(b) Critical analysis of triple parentage

The main facet of the triple-parent argument is that step-children would be unfairly advantaged by an additional intestacy claim. A criticism of this argument is that the legislation need not fully equate step-children to biological children. Although a step-child may have three or more parental intestacy claims, any claims against step-parents could be reduced in comparison to their claims against a natural parent's estate.

There are ways to limit the claim of a step-child to ensure that the biological child maintains a better position. For example, step-children's rights need not extend to protection from disinheritance. Their right in succession could only apply where no valid will existed and the estate fell to intestacy. Another solution would be to limit the value of a step-child's right to a percentage of what a biological child would be entitled to, something suggested by Sparks in relation to cohabitants' rights.[86] Thus, the value of a step-child's claim could be reduced to ensure that they are not overly advantaged. There will be further deliberation on the appropriate formation of step-children's rights later in the discussion.[87]

[83] Reid (n 67) at 397.
[84] K McK Norrie, "Reforming Succession Law: Intestate Succession" (2008) 12(1) Edin LR 77 at 79.
[85] *Ibid.*
[86] E Sparks, "Changes to the Laws of Succession – The Importance for Family Lawyers" (2017) Fam LB 1 at 4.
[87] See below, generally Section E.

A further criticism is that every person in Scotland has a varying number of siblings, cousins, aunts, uncles and children from whom they can inherit. It therefore seems overly critical to use the number of potential parental estates a person can claim from as a valid reason to reject step-children's claims. In addition, some women are unaware of the identity of the father of their child resulting in no father being registered on the birth certificate. The fact that these children are only able to claim intestacy rights on one estate has not been considered unfair as it is an accepted family situation in modern society.

The triple-parent argument is most effective as an argument against step-children's rights in protection from disinheritance as this would involve the overriding of a testator's specific intentions. However, for general intestacy, the triple parent argument is less persuasive as there are ways to formulate step-children's intestacy rights to avoid unfairly advantaging them. Furthermore, it is important to acknowledge that legislation cannot remove all the inconsistencies that arise from a family's size and background. The rules must attempt to apply to the majority of family situations but it would be unrealistic to expect the intestacy rules to perfectly suit every person in Scotland whilst also incorporating the guiding principles.

(4) Evaluation

Although the SLC and academics in this area agree that step-children's rights are not necessary in the law of intestate succession, the view of the present writer is that there is a place for additional rights. The main argument refuting step-children's intestacy rights posited by the SLC is in relation to triple parentage. However, the advantages felt by step-children could be reduced by the way the law is drafted. Furthermore, the reliance on arguments involving adoption and will-making is unfounded because, when considered in more depth, they are not as simple as Kenneth Reid and the SLC suggested.

E. PROPOSAL FOR STEP-CHILDREN'S INTESTACY

To analyse possible reform options available for Scots law, there must be a clear definition of who would qualify for step-children's intestacy rights. This section will attempt to define a step-child before looking at the approaches taken in different jurisdictions and whether they would be appropriate in Scotland. Attention will then be paid to the potential ways of drafting the law, particularly looking at ensuring compatibility with our guiding principles.

(1) Definition of a step-child

A step-child is defined by the Oxford Dictionary as a "child of one's husband or wife by a previous marriage".[88] Yet, there is no definition of what a step-child is for succession law. This may be because the dictionary definition is currently sufficient for legal purposes. Within this definition, there can be a further distinction made between dependant step-children and adult step-children. When forming a system that grants intestacy rights to step-children, there are different definitions of a step-child that could be considered. The strictest definition would only include dependant step-children as they are considered the worthiest. The broadest definition of a step-child would include the step-children of cohabitants as this goes beyond the scope of the dictionary definition. This section will determine the appropriate definition of a step-child in the potential reform of Scots law.

(a) The worthy step-child

Unlike the equal treatment of natural children, there is an argument to be made that one category of step-children is more *worthy* than another. If you assume that a *worthy* child is a step-child through

[88] R E Allan (ed), *The Concise Oxford Dictionary of Current English*, 8th edn (1990) 1195.

marriage or civil partnership who has been accepted by the deceased, you could draft a provision that applies only to them. The benefit of this approach is that these step-children are the closest in definition to natural children and there is less chance of public criticism. The SLC acknowledged the practical difficulties of acceptance being established and noted the concern of allowing a step-child to inherit from their step-parent's extended family.[89]

In relation to the guiding principles, applying intestacy provisions to *worthy* step-children would not be considered administratively efficient. In practice, this definition is not objective and would result in a court having to determine whether applicants were sufficiently accepted by the deceased. It also falls short on the simplicity principle as it is difficult to define acceptance as the criteria is not based on a blood relationship. It would, however, satisfy the fairness of distribution principle. The only step-children entitled to a claim would be those who would have been included in their step-parent's testament.

(b) The strict definition

A strict definition of a step-child would have the benefit of retaining the simplicity of our current system, although this would be done at the expense of the fairness of distribution principle. In Queensland, a strict definition applies to step-children under Section 40A of the Succession Act 1981.[90] A step-child is limited to the "child of a spouse of the deceased person",[91] but does not extend to the child of a cohabitant. Furthermore, the provisions do not apply to a step-child when their step-parent and natural parent have divorced.[92]

A strict definition in Scots law could limit intestacy rights to step-children through marriage who lived with their step-parent for a minimum period of time. This definition would mirror the provisions of Queensland by excluding children of cohabitants and children whose step-parent and natural parent have divorced.

This definition satisfies the simplicity and efficiency principles as it is a fixed objective standard and will be less likely to result in litigation. Although the fairness of distribution principle is not the paramount consideration here, an attempt has been made to take it into account by limiting the scope of the intestacy claimants by introducing a time period. This has the effect of extending rights to step-children who are more likely to have had a meaningful relationship with their step-parent but it reduces the practical difficulties of litigation that would arise if you required acceptance of the child by the deceased.

(c) The broad definition

Cohabitants are becoming more common in Scottish society and their rights are beginning to be recognised. It could be argued that the definition of a step-child should be extended to include the child of a cohabitant and ought not to be restricted to relationships formed through marriage and civil partnership.

In England and Wales, step-children are not heirs on intestacy but children who "in relation to any marriage" were treated by the deceased as a "child of the family" have the right to apply to the court for an order.[93] The Law Commission recognised that this marriage requirement was a reflection on the fact that financial provision can be made for a "child of the family" on divorce .[94] In the 2011 Report, it was recommended that this claim be extended to a "child of the family" where there is no

[89] Report on Succession (Scot Law Com No 215, 2009) para 2.32.
[90] Queensland Succession Act 1981, s 40A.
[91] Queensland Succession Act 1981, s 40A(1).
[92] Queensland Succession Act 1981, s 40A(2).
[93] Inheritance (Provision for Family and Dependants) Act 1975 s 1(1)(d).
[94] Intestacy and Family Provision Claims on Death (Law Com No 331, 2011) para 6.29.

underlying marriage. The Law Commission regarded the provision in the 1975 Act as being "outdated" and stated that there was no reason for a cohabitant with a similar quality of relationship to be excluded as a result of the wording of the section.[95]

It is likely that intestacy reform extending beyond the traditional definition of a step-child to the children of cohabitants would be premature in Scots law as cohabitants' rights and step-children's rights are not yet solidified. Additionally, the SLC values certainty and such a provision would partially remove this from our system. However, it is a possible expansion that should be kept in mind for future reform.

(d) Evaluation

The strict definition is the most applicable to our system where a step-child's relationship is established through marriage and their entitlement is established where they have resided with their step-parent for a fixed period of time. Although a broader definition would be the most progressive approach, the S(S)A 2016 implied that the nature of reform in this area is incremental and such a definition may be too far ahead of its time.[96] The time period element of the strict definition would address the fairness of distribution principle and would prevent a natural child's claim being disproportionately affected by a step-child who had little connection to the deceased. It could also extend to adult children as potential claimants where they satisfied the time period, unlike a system aimed solely at dependants. Furthermore, a fixed definition is the most suitable because a discretionary system would not comply with the administrative efficiency principle and is not in tune with the general approach taken in Scottish succession law.

(2) Discretionary rights for step-children

(a) General

Scots law has traditionally used fixed rights as a means of dividing an estate in intestate succession.[97] This is in line with the guiding principles as discretion tends to lead to administrative inefficiency. However, a discretionary approach has been taken in relation to cohabitants in Scots law and several other jurisdictions use a discretionary system in relation to step-children's intestacy rights.

(b) Other jurisdictions

(i) England and Wales

As discussed previously,[98] England and Wales recognises step-children's rights on intestacy in the Inheritance (Provision for Family and Dependants) Act 1975. These rights are discretionary, and an order allows a court to award "reasonable financial provision",[99] which, in relation to step-children, is defined as a provision for maintenance. There is no limit on a step-child's claim and they can be awarded any sum which the court deems reasonable.

This system is more flexible than the Scottish system for cohabitants' claims. Scotland could also impose a limit to the amount a court can award if a discretionary system of this type were to be adopted and this would prevent step-children from being awarded a claim worth more than a natural child in the same position

.

[95] *Ibid.* para 6.40.
[96] See generally Sections A and B for discussion of S(S)A 2016 and the pace of reform.
[97] See generally, Succession (Scotland) Act 1964.
[98] See above Section D(1)(c).
[99] Inheritance (Provision for Family and Dependants) Act 1975 s 2(b).

(ii) Western Australia

In 2013, the Family Provision Act 1972 added a provision to allow step-children to apply for a claim on the deceased's estate.[100] These rights apply to step-children who were being maintained or were entitled to maintenance prior to the death of their step-parent, which is similar to the dependency claim that is available in England and Wales.[101] The Act goes further to address the second spouse issue previously discussed.[102] Under Section 7(1)(eb), a step-child is entitled to a discretionary claim where more than a prescribed amount of their step-parent's wealth came from their natural parent.

(c) Inspiration from cohabitants' rights

Inspiration could also be drawn from our domestic law of cohabitation. As is the case with step-children, cohabitants do not have a blood relationship with the deceased or have a relationship that is the legal equivalent to a blood relationship. Step-children's rights could therefore be formulated in way that is akin to cohabitants' rights, ensuring consistency within the law.

At present, cohabitants can apply to the court for a discretionary claim under the FL(S)A 2006; however, there are restrictions on these rights which are intended to protect the position of spouses. The main restriction is found in Section 29(4) of the Act, which states that an order:

> "shall not have the effect of awarding to the survivor an amount which would exceed the amount to which the survivor would have been entitled had the survivor been the spouse or civil partner of the deceased."[103]

This provision prevents a cohabitant's discretionary right from being valued more than a spouse in the same position. A similar restriction could be applied to step-children's rights where their claim on intestacy could not exceed that of a natural child. This would limit the triple parentage argument as step-children would not be as unfairly advantaged. However, this provision only prevents a cohabitant's right from exceeding that of a spouse and does not prevent their claims from being equal in value. It is probable that the SLC would consider this to be inadequate as equality with biological children has previously been considered too great a change.

(3) Fixed rights for step-children

(a) General

A fixed system has the benefit of simplicity as all qualifying step-children would be entitled to the same claim. There are numerous factors that a fixed system would have to take into consideration. For example, the fair distribution principle must be adhered to in order to prevent natural children from being significantly disadvantaged as a result of reform. To minimise the effects of the triple parentage argument, an attempt should be made to ensure that natural children are prioritised in the estates of their natural parents. This could be achieved by the value of a step-child's claim being smaller or the order of succession prioritising a natural child.

[100] Western Australia Family Provision Act 1972 s 7(1).
[101] Western Australia Family Provision Act 1972 s 7(1)(ea).
[102] For this discussion see Section D(1)(b) above.
[103] Family Law Scotland Act 2006 s 29(4).

(b) Other jurisdictions

(i) China

China recognises step-children as fixed heirs with their rights dating back to 1985.[104] Under Article 10,[105] children are listed as heirs and inherit after spouses but before parents. Furthermore, all "children" are treated equally. The Article goes on to define "children" as:

> legitimate children, illegitimate children and adopted children, as well as step-children who supported or were supported by the decedent.[106]

Within this system, a step-child's worthiness is derived from their dependency on their step-parent or the support they provided to their step-parent. These factors are indicative of a meaningful relationship between the deceased and the step-child, therefore reducing the likelihood of an unknown step-child being entitled to a claim. This form of a step-child's right is appropriate as it only applies to step-relationships where there has been support or maintenance. However, with regard to its application in Scots law, the granting of step-children's rights equal to that of natural children is not appropriate in accordance with our fairness of distribution principle.[107]

(ii) California

In 1983, the California Probate Code introduced a narrow two-part test that allows step-children to succeed on intestacy.[108] It permits succession through a foster parent or step-parent where:

> (a) the relationship began during the person's minority and continued throughout the joint lifetimes of the person and the person's foster parent or step-parent.
> (b) It is established by clear and convincing evidence that the foster parent or step-parent would have adopted the person but for a legal barrier.[109]

Where a step-relationship is established, a step-parent and step-child are given intestacy rights equal to that of a parent and child. However, there are some problems with the way these provisions have been drafted. Primarily, it is difficult to prove the deceased's intentions in their attempt to adopt the step-child. Moreover, for the second condition to be satisfied, the legal barrier to adoption must have been continuing for the length of the step-relationship. One admirable aspect of the California Probate Code is that it acknowledges that adoption is not an easy solution to intestacy issues.

Although this relationship is stricter than the approach used in China, it is arguably still not appropriate for Scots law. This is because step-children and natural children would be given equal rights, and this would not satisfy the fairness of distribution guiding principle.

(c) Section 2 right

The least controversial approach to this issue would be to insert a step-child's right into Section 2 of the S(S)A 1964. Scotland has a system of unlimited heirs that has the potential of benefitting "laughing heirs".[110] Step-children are arguably more likely to have had a meaningful relationship with the deceased. Therefore, a provision in favour of stepchildren could be inserted between subsection 2(1)(h) and subsection 2(1)(i). This would mean that the only beneficiaries disadvantaged by such

[104] Law of Succession of the People's Republic of China 1985.
[105] Law of Succession of the People's Republic of China 1985 Article 10.
[106] Law of Succession of the People's Republic of China 1985 Article 10.
[107] For discussion of this principle see above at section B(1)(a)(iii).
[108] California Probate Code s 6454.
[109] California Probate Code s 6454.
[110] K Anderson, "Intestacy in Scotland: The Laughing Heir" (2011) 2(1) Aberdeen Student L Rev 52.

intestacy rights are the "ancestors of the intestate" that are "remoter than grandparents".[111] This is the approach suggested by the 2008 revisions to the Uniform Probate Code, which states that a step-child can succeed after other relatives but before the State.[112]

The benefit of this approach is that there are only likely to be a small number of heirs who would be opposed to these rights being implemented. Only "laughing heirs" would have their rights weakened. Furthermore, "laughing heirs" would not have a strong claim against step-children's intestacy rights as they are unlikely to have known the deceased on a meaningful level. This system would adhere to the guiding principles as it would be a simple amendment to our law.[113] In addition, it would be administratively efficient as a fixed system does not invite litigation. Moreover, the intestacy rights of current beneficiaries would remain largely unaffected.[114]

The main criticism of this approach is that it does not sufficiently improve the position of a step-child and cannot justify a change to the law at all. However, it appears to be the only approach that would be approved by the SLC and accepted by the public. Furthermore, it could be argued that slight improvement is better than no improvement as there will be some circumstances when step-children would be able to claim where they previously could not.

(d) Fixed percentage of a natural child's entitlement

Cohabitants' rights were considered when reviewing possible discretionary approaches to step-children's intestacy rights. They are also worth considering in terms of fixed rights. An alternative restriction on step-children's rights could be to limit their claim to a percentage of what they would have received had they been a natural child. This formation was discussed by Sparks in relation to cohabitants.[115] The SLC had proposed a discretionary award for cohabitants where they would be entitled to a percentage of what a spouse would be entitled to.[116] They stated the percentage should be discretionary and that it should depend on the period of cohabitation, the interdependence of the parties, and the contribution made by the survivor to their shared life.[117] This approach makes the size of the estate and the identities of other beneficiaries irrelevant.[118]

Although, this formation of a claim is discretionary, the general principle of subjecting a cohabitant's claim to a percentage reduction could be applied in a fixed way.[119] It could be possible to award step-children a claim under Section 2 of the S(S)A 1964, but to apply a percentage reduction in scenarios where they would be able to claim from three respective parents.

(4) Evaluation

Discretionary rights are more in line with the fairness of distribution principle and have the benefit of being able to award deserving claimants with the appropriate sum. Furthermore, this approach has been adopted for cohabitants in Scots law. However, cohabitants' discretionary rights have been criticised by the SLC.[120] This suggests that a discretionary system may not be appropriate for step-children in Scots law.

Fixed rights are simpler and conform to the traditional form of intestacy law in Scotland. Nevertheless, a fixed system is less able to cater to the individual needs of modern families. Although

[111] Succession (Scotland) Act 1964 s 2(1)(i).
[112] United States of America Uniform Probate Code 2008 Revision s 2-103(b).
[113] See section B above.
[114] See section C above.
[115] E Sparks, "Changes to the Laws of Succession – The Importance for Family Lawyers" (2017) Fam LB 1.
[116] Report on Succession (Scot Law Com No 215, 2009) para 4.14.
[117] *Ibid.*
[118] Sparks (n 115) at 4.
[119] Although the discretionary formation is discussed below.
[120] Report on Succession (Scot Law Com No 215, 2009) para 4.8.

a fixed system cannot cater for all eventualities, inspiration could be drawn from the Chinese system by drafting the legislation in a way that better recognises the rights of *worthy* claimants. Although this has the effect of removing some certainty from the law, it could be the most suitable way forward in Scotland.

Of the recommendations identified, the most suitable approach is a section 2 intestacy right placing step-children before remote ancestors in the order of succession. Although this may only benefit step-children where there are no closer claimants, the SLC has been reluctant to change the law and this may be the broadest right they would be willing to adopt.

F. CONCLUSION

The purpose of this article has been to examine the current law of intestacy in relation to step-children, and to question whether it is suitable for modern society. Step-children should be given fixed rights under Section 2 of the S(S)A 1964. One option could be to grant them intestacy rights prior to remote ancestors in subsection 2(1)(i). The benefit of this scheme would be that step-children are likely to have a more meaningful relationship with the deceased than a "laughing heir". This is desirable as the proximity of the relationship is an important factor considered by the SLC when drafting reform proposals. Moreover, under this system the introduction of step-children's intestacy rights would not impact the current rights of other claimants as they would take priority under Section 2. Although the position of a step-child would only be marginally improved if this system were adopted, it is argued that such a change is better than no change.

SHOULD PARTIES BE ABLE TO CREATE REAL RIGHTS OF THEIR OWN DEVISING?

Susanna Mulvihill[*]

A. INTRODUCTION

The principle of *numerus clausus* is "the idea that both the number and content of property rights is limited",[1] as opposed to the potentially unlimited number of personal rights that may be created by contract between parties. Property law is mandatory law, with effects against third parties, so there is much to be said for restricting the number of real rights available. However, if in Scotland we base these rights (albeit loosely) on the same list used by the Romans two thousand years ago, are we not being rather belligerent by refusing to allow parties to devise new rights that may be more appropriate to twenty-first century life? I shall argue in this article that, in certain circumstances and under the appropriate conditions, parties should be allowed to create their own real rights in relation to heritable property, provided there are registration facilities available to satisfy the need for publicity in property transactions.

The debate surrounding *numerus clausus* has attracted much international consideration, but I shall be predominantly focusing on the impact that permitting parties to create their own real rights may have in Scotland. In Section 'A' I begin by examining the concept of *numerus clausus* and the benefits a closed list may bring, both commercially and economically. I argue that in some circumstances opposition to expanding the fixed list can be easily overcome through registration. Then I consider whether the doctrine of *numerus clausus* truly applies in Scotland, given that academic literature acknowledges the accepted list of real rights cannot be said to be definitive. To demonstrate an alternative approach I look to the South African model, considering whether a similar attitude might be desirable in Scotland. I conclude this section by stressing the importance of public registration should the law admit parties to create their own property rights.

It would be trite to recount the details of Jack and the Beanstalk here,[2] but in the second half of this article I reimagine the fairytale's ending to illustrate three different scenarios in which parties potentially should be able to create their own property rights by agreement: through personal servitudes, commercial leases, and over moveable property. I argue that in the third scenario a *numerus clausus* approach is preferable, but in the first two cases there is a strong argument for allowing parties to devise rights of their own choosing.

---

[*] 3rd year LLB student at the University of Edinburgh.

[1] B Akkermans, "The Numerus Clausus of Property Rights" in M Graziadei and L Smith (eds), *Comparative Property Law: Global Perspectives* (2016).

[2] But for a classic retelling see V Southgate, *Well-Loved Tales: Jack and the Beanstalk* (1971).

Overall, I conclude that Scotland cannot be said to have a definitive fixed list of real rights, and in certain circumstances, such as commercial leases and personal servitudes, parties should be able to create their own real rights over heritable property, provided there are adequate registration facilities available to fulfil the requirement of publicity. However, for corporeal moveable property the existing regime should remain to prevent uncertainty and confusion.

B. *NUMERUS CLAUSUS* IN SCOTLAND

(1) Why have a fixed list?

For its proponents, the idea of having a fixed, closed list of potential real rights is beneficial in part because property law is "mandatory law" which governs how rights can be created, altered, and discharged.[3] As real rights have third party effect there must be an element of universal certainty for any person who may be affected by them. As such, the argument follows that there is a "need to establish a basic set of entitlements".[4] It is contended that if parties could create their own real rights it would potentially lead to an excessive fragmentation of property rights,[5] resulting in the devaluation of property, a reduction in the benefits of trade, and difficulty in keeping track of what rights affect any property. It has been said that the doctrine of *numerus clausus* provides freedom within ownership as it limits the rights others may have over property. It has also been said to promote the free circulation of goods, because those dealing in property have a degree of certainty as to what rights may affect it. Further, it has been purported that to allow parties to create their own real rights would be to blur the distinction between real and personal rights. On the other hand, it has also been noted that the "adoption of closed systems […] simultaneously stifles legal development and encumbers adaptation to modern conditions",[6] so it is necessary to assess the respective merits of both positions.

One of the main arguments against permitting parties to create their own real rights is highlighted by Henry Smith's example of a so-called "Monday-right".[7] In brief, Smith describes a scenario in which a party who is the absolute owner of a watch wishes to sell a right in it to another party, the content of the right being ownership on a Monday. One of the issues noted with this sort of arrangement is that it leads to costs – not just for the parties to the timeshare, or their successors, but also to the world at large. The reason for this is that anyone wishing to transact with any watch has the additional burden of trying to ascertain whether there is a similar right affecting it. These third-party information costs would affect the market by raising the cost of transacting overall and affect the alienability of property.

I concede that these are important, and persuasive, arguments against allowing parties to create their own real rights for corporeal moveable property; arguments that tip the scale towards the *numerus clausus* principle. However, with heritable property the situation is different. This is in part because heritable property (or ownership of it at least) must be registered, alongside any burdens on that property. Anyone transacting with heritable property already undertakes searches to determine what burdensome conditions affect it, or at least they should if they have at least a modicum of wit. If the new rights affecting the property are registered, they would be discoverable, at no additional cost, to those already in existence. It could be argued that such rights would potentially clutter the Land Register, making searches more cumbersome and time consuming, but there is little evidence to

[3] Akkermans (n 1) at 7.
[4] H E Smith, "Standardization in Property Law", in K M Ayotte and H E Smith (eds), *Research Handbook on the Economics of Property Law* (2011) 148.
[5] *Ibid*. at 148; M A Heller "The Boundaries of Private Property" (1999) 108 Yale LJ 1163 at 1176-1178; F Parisi, "Entropy in Property" (2002) 50 American Journal of Comparative Law 595.
[6] C G Van der Merwe, *The Law of Things* (1987) para 8.
[7] Smith (n 4).

support this claim, and, as argued below, it is unlikely that the rights affecting property would exponentially grow to the extent that cluttering would be the result.

Another argument against allowing a time-share arrangement with heritable property is that it would have the effect of creating multiple ownership rights in the same property, a concept horrific to Scots lawyers. I contend that (a) at any one time there would only be one right in existence, although that right may periodically flip between parties, and (b) allowing timeshares to be real rights would be a good thing in general. It would mean that people could have a more determinate right in properties they have an interest in, as currently such arrangements must be constituted by trusts, wherein the parties have merely a personal right. Moreover, such trust arrangements are opaque when trying to ascertain true, beneficial property ownership, and this is a hindrance to the principle of publicity in Scots law. As such, for heritable property at least, I must respectfully disagree with Henry Smith.

It has also been claimed that permitting parties to create their own real rights could potentially produce "untrammeled expansion which leads to obscurity".[8] This argument has more than a whiff of the 'slippery slope' fallacy about it. What it fails to acknowledge is that parties to a transaction ultimately have their own interests at heart and are unlikely to do much to damage them. If A is selling property to B and wishes to create a right, he has little reason to create a right that is potentially going to lead to a significant decrease in the price he can demand for it. If B is purchasing, it is improbable he is going to agree to conditions that may drastically reduce the value of the property. Moreover, neither party is likely to want to constitute too onerous a right on the property, as others may follow their lead and they may end up just as frustratingly burdened themselves in future transactions.

In respect to the argument that allowing parties to create their own real rights could lead to a blurring of lines between real and personal rights, it should be noted that in some circumstances personal rights can already impose a burden upon the owner of property.[9] For example, there may be a contractual prohibition on alienation of land, stipulations in a trust deed, rights of pre-emption, etc. In Scotland, the "off-side goals" rule even permits a personal right to "trump" a real right in specific circumstances.[10] Furthermore, even the institutional writers acknowledged that the owner of property may use and dispose of property, except where "restrained by law or paction".[11] This recognition indicates that a real right, even in the historic roots of modern Scots law, does not necessarily lead to certainty for the wider world not party to a contract. True, the examples above involve burdens on the owner of a property that are personal in nature. However, they do relate to a limited personal right another party may have over property. Is it really fair to allow the owner to confound their obligation to another merely by disposing of the property?

(2) The fixed list in Scotland

In Scotland the officially recognised list of real rights is as follows: ownership, rights in security, proper liferent, servitude, lease, possession, public rights, and exclusive privilege. Although it is claimed that this is a fixed list,[12] "the possibility that other real rights exist cannot be entirely excluded".[13] This leads to an interesting conundrum. Can it really be said that Scotland has a closed list if we cannot find the door? If there is a possibility that other real rights "as yet unrecognised or undiscovered" may be added to the list in future,[14] surely it is little more than a charade to claim that

[8] R Paisley, "Real Rights: Practical Problems and Dogmatic Rigidity" (2005) 9 Edin LR 267 at 281.
[9] Van der Merwe (n 6).
[10] *Rodger (Builders) Ltd v Fawdry* 1950 SC 483.
[11] Erskine *Inst* II.i.1.
[12] K G C Reid, "UK: Scotland Real Property Law", available at *https://www.eui.eu/Documents/DepartmentsCentres/Law/ResearchTeaching/ResearchThemes/EuropeanPrivateLaw/RealPropertyProject/Scotland.PDF*.
[13] *The Laws of Scotland: Stair Memorial Encyclopedia* vol 18 para 5, footnote 1.
[14] R Paisley (n 8) at 269.

the list is truly fixed. Courts have been known to 'discover' rights that are part of the *numerus clausus* of real rights in Scotland,[15] but this is merely extending the fiction that the list is, in fact, fixed.

The list may also be, and has been, added to by statute, meaning that expansion is possible and therefore cannot be said to be as determinate as claimed.[16] Even at common law, the list must be able to be said to have been added to by the emergence of real burdens in the early nineteenth century.[17] If the list has required supplementation to fulfil the needs of modern society, then it is demonstrative that the fixed list is inadequate. Legislatures, however, are generally reactive when assessing modern needs. They merely respond once a problem has crystallised. By allowing parties, in certain circumstances, to create their own real rights it would minimise the possibility of stagnation in the economy and allow the law to evolve with more fluidity.

Even jurisdictions that are strict adherents to the *numerus clausus* principle have needed to concede a certain amount of flexibility.[18] Moreover, by removing the requirement that legislation is necessary before the list may be expanded, there would be a reduction in the costs associated with passing legislation. This would be in the public interest. In addition, problems have arisen in the past with poor quality of legislative drafting, for instance with the Gas Act 1986.[19] Here the Act gives the power to gas transporters to open streets and lay pipes, but the drafters omitted to give the right to the transporters to use the pipes after they had been laid. Such a significant oversight is indicative of the dangers of putting the responsibility of defining boundaries of rights into the hands of the bureaucrats.

It has been said that although the fixed list of available real rights in Scotland is limited, there is the ability to expand the scope of existing rights to fit more modern needs. However, this is normally done through judicial recognition, and although the legislature can be slow to respond to modern needs, compared with the judiciary they are the Olympic sprinters of legal development. In the three hundred years since the list of recognised servitudes was famously increased through judicial recognition there have been few further advancements,[20] and the recent case of *Moncrieff v Jamieson*[21] (although significant) can hardly be said to be revolutionary when faced with the needs of modern life. Therefore, those best placed to expand the list of real rights are those who are using them; namely the parties, and as such should, in principle, be able to create their own.

(3) The South African model

Unlike Scotland, South Africa does not recognise a fixed list of real rights in relation to heritable property. This is in part because "the demands of modern commerce and legal practice render it necessary sometimes to create rights that do not resemble the typical Roman Law kinds".[22] However, it is in no means a model which allows parties to create their own real monsters and release them into the world. The real rights must be registrable,[23] and as such can only apply to heritable property. The requirement of registration means that any concerns about the lack of publicity, or obscurity, of the real right in question are allayed. This is because the newly determined right will be readily apparent on a search of the register and supports my earlier contention that permitting new rights to be registered should not make such a search more cumbersome or time consuming.

---

[15] For instance, *Bowers v Kennedy* 2000 SC 555.
[16] For example, the occupancy rights provided by the Matrimonial Homes (Family Protection) (Scotland) Act 1981.
[17] *Tailors of Aberdeen v Coutts* (1840) 1 Rob 296, although there is debate as to whether real burdens are actually real rights or real obligations, but that is beyond the scope of this essay: See K G C Reid, "Real Rights and Real Obligations" in S Bartels and M Milo (eds), *Contents of Real Rights* (2004).
[18] For instance: the German principle of "Auflassungsanwartschaftsrecht".
[19] Gas Act 1986 Sch 4 para 147.
[20] *Town of Falkirk v Carmichael* (1708) Mor 10916.
[21] *Moncrieff v Jamieson* 2008 SC (HL) 1.
[22] H Mostert and A Pope (eds) *The Principles of the Law of Property in South Africa* (2010) para 3.2.2.2.
[23] Deeds Registries Act 47 of 1937 s 3(1).

For a new right to be registered, two more elements must be present for it to be recognised as 'real'. Firstly, there must be a clear intention of the parties to bind their successors,[24] and secondly, the right must amount to a "subtraction from the dominium".[25] This subtraction from the dominium test means that the right must constitute a hindrance to, or reduction of, the ownership right of the property. In practice there have been very few rights created, and the courts have been strict when determining which rights may be recognised as real. So, when testing the claim that permitting parties to create their own real rights would create obscurity, leading to some sort of exponential, chaotic expansion, strongly detrimental to the law of property in Scotland, it can be demonstrated through South African practice that this is a concern without foundation, provided there are adequate safeguards in place.

Primarily, and this does much to counter the claims of proponents of a *numerus clausus* of real rights, the courts have been diligent in their assertion that "if a right is personal in nature it cannot be transformed into a real right by the intention of the parties".[26] In the case of *Lorentz v Melle*[27] it was held that an obligation to make monetary payment was a personal obligation, specific to the party upon which it fell, and was not an obligation that could be attached to the land. Below I will consider whether in fact certain personal obligations should be permitted to be made real, but in the South African model it provides an example where, so long as the boundaries in which parties may create their own real rights are tightly drawn, *numerus clausus* is an unnecessary restriction in the twenty-first century.

C. JACK, THE GIANT, AND PERSONAL SERVITUDES

So far, I have concentrated on the general theoretical arguments for the continuance of a *numerus clausus* of property rights, and indicated why those claims may not hold the weight generally attributed to them. In this, and the following sections, I shall now entertain three instances in which it may be beneficial to allow parties to create their own property rights. Using Jack and the Beanstalk as a starting point, I shall reimagine the ending of the tale, then apply the three situations to possible real-world scenarios. To begin with, in this section I shall examine the possibility of parties being able to create their own personal servitudes. In Section 'D' I shall look at commercial leases and their effect on singular successors. Finally, in Section 'E', I shall consider whether parties should be able to create their own real rights in relation to corporeal moveable property.

Imagine, if you will, that instead of Jack being a thief and cutting down the beanstalk, he and the Giant became friends. They entered into a commercial arrangement involving the singing harp and the golden goose, and Jack visited the Giant regularly. With his newly acquired wealth he sold his farmhouse and moved into a luxury apartment block some distance away, not neighbouring his former abode. The beanstalk was sold with the farmhouse, as Jack didn't want the cost of maintaining it, nor any problems that may have arisen from falling foul of the High Hedges (Scotland) Act 2013. However, Jack wished to retain a right of access to the Giant's property via the beanstalk for himself and his family. As such, when disposing of the farmhouse he inserted a clause in the missives retaining a perpetual right for him and his successors to use the beanstalk.[28] If this right is entered in the burdens section on the title sheet, clearly identifying Jack, is there any good reason this should not be binding on subsequent acquirers of the farmhouse?

[24] *Nel NO v Commissioner for Inland Revenue* 1960 (1) SA 227 (A).
[25] *Schwedhelm v Hauman* 1947 (1) SA 127 (EDL) 135.
[26] Van der Merwe, (n 6) para 49.
[27] *Lorentz v Melle and Others* 1978 (3) SA 1044 (T).
[28] As the beanstalk is adjacent to the farmhouse, it is unlikely that Jack would be able to resort to the Land Reform (Scotland) Act 2003 s 6. See also *Gloag v Perth and Kinross Council* 2007 SCLR 530.

The idea of the retention of a personal right to use property arose in the nineteenth century case of *Scott v Howard*.[29] In brief, the original disponers of a theatre retained both an annuity of £2 per annum and the right to enter the theatre to watch any performance they chose. The theatre subsequently burned down twice and was sold again. The House of Lords found that the annuity payment was connected to the land and was therefore binding on successors. However, the right to enter the theatre was not: it was a personal right and not able to be made real under the law of Scotland.[30] Because of the limited nature of real rights in Scotland, the parties who made the original agreement were unable to fully realise the terms they stipulated in their contract. True, prior to the original disposition the owners were in a significant amount of debt, which the purchaser undertook to pay in lieu of price for the property, but still, without the clause regarding theatre admission the original disponers may have demanded either a higher annuity or an additional price. Moreover, in the subsequent disposition the terms of the original agreement were referred to (and agreed to), so it could be argued that the original disponers were unfairly prejudiced by the later judgement.

This ultimately boils down to an ideological question: should priority be given to the sellers' right to determine certain stipulations as to how the property is to be used, or to the ultimate right of the purchasers to use their property unfettered by the ghostly hand of owners' past. One could argue that when the property was previously owned the proprietors at that time had a right to determine its conditions of use and could exercise that right even when disposing of the property. On the other hand, why should the new acquirers have an inferior right dictated by people who no longer have an interest in the property, save by some invention regarding that which they wished to retain? Ultimately, there is probably no satisfactory answer to this question, save that the law should protect those in current possession over those who have willingly surrendered it. However, if the new acquirers have agreed to the transfer in full knowledge of the conditions affecting the property, then there is something deeply unsatisfactory about permitting them to disregard those conditions.

As the law currently stands, a servitude must be praedial;[31] that is, it must be for the benefit of a property. However, a property is not sentient: it cannot enforce its rights on its own. The right of servitude can only be enforced by people, people who gain the benefit of the right through their association with the property. Thus, the benefit is not in truth for the property, so why should there not be provision for servitudes to be directly constituted for individuals? The answer lies in Roman law,[32] where servitudes are defined as belonging to land. The question then arises as to whether it is wholly appropriate to cling to these ancient definitions in the modern age. In my opinion, it is not.

In the case of *Scott v Howard* mentioned above it should be noted that the court took into account that the theatre the original disponers were trying to gain access to was materially different, as it had been rebuilt twice since the original sale. Therefore, this made the access provision harder to justify. But what if it had been the same theatre? Or what if, instead of a theatre, the sellers had been disposing of a piece of undeveloped (and undevelopable) land and were part of a naked rambling association who still wished to occasionally use it for their activities? This may seem like a flippant and explicit example, but it highlights the fact that the attempt to create a personal servitude would only be likely in highly specific circumstances for highly specific reasons. These types of unique circumstances are unlikely to be provided for by expanding the current list of servitudes. To rely on the goodwill of singular successors to honour the agreement made by their predecessors is a risky gambit. Therefore, the law should in theory permit parties to create personal servitudes.

The question then arises as to what restrictions should be placed upon the creation of personal servitudes? In the South African model mentioned above, new real rights cannot be personal in nature, so we cannot adopt the path taken there. Certainly, it would be necessary for the right to be registered

[29] *Scott v Howard* (1881) 6 AC 295.
[30] *Ibid*, at 312 per Lord Watson.
[31] Title Conditions (Scotland) Act 2003 s 75. NB although not stipulated, the praedial nature is implied by the need to register a servitude against both a burdened and a benefitted property.
[32] P Birks and G McLeod (trans) J *Inst* 2.2.3.

in the Land Register, so potential successors are aware of it before they agree to purchase the land. The parties to whom it applies should be clearly identified, so the right is not open to being exploited by the world at large to the detriment of the owner. If the parties 'disappear' the owner should be able to apply to the Keeper to have the personal servitude removed. The right should be a right to enter and use the land for a limited and well-defined purpose, again so the inhibition on ownership is not too onerous. Finally, and this would perhaps be disappointing to the naked ramblers, the use should not be contrary to public policy or the public good. Any request to create a personal servitude should be formally made to the Keeper, the costs of which should be borne solely by the party attempting to invent it. The public policy criteria may seem like an ambiguous restriction but permits a certain amount of discretion to the Keeper to reject inappropriate applications, while still leaving enough scope for parties to tailor servitudes to their needs.

D. COMMERCIAL LEASES

Now I return to Jack and his beanstalk. Prior to selling the farmhouse and moving to his luxury apartment, environmentally conscious Jack granted a lease to a renewable energy company which contained an option to develop wind turbines on the beanstalk. Following the sale, the new owners decided such machinery would be unsightly and contested the option. The option, not being *inter naturalia* of a commercial lease,[33] was found not to be binding upon the successors to the heritable title. In this section I consider whether it, and other clauses in the contract, should have been.

In a recent address to the WS Property Law Panel,[34] Scottish Law Commissioner Caroline Drummond indicated that the Scottish Law Commission is considering whether to consult on this question.[35] This is partly because of the benefits it would offer to the commercial property sector. It is common practice for proprietors of commercial property to grant long leases to prospective purchasers instead of conveying the property.[36] If the proprietors then convey the property to a third party, it can place the tenant in a precarious position, especially if the lease was for a specific future purpose which is not "normal" in a lease. To allow conditions in a lease to be binding on successors would be to give the whole lease real effect; that is, the terms of the lease would be good against the world, or at least against anyone succeeding to the lease or ownership of the land. Although most problems with this arise in the context of purchase options, it is equally applicable to circumstances where there is an option to develop renewable energy plants.

It is unnecessary to recite the reasons for encouraging the development of clean, alternative energy sources wherever possible. Nor is this article the proper forum for debating the merits of such a policy. However, I shall proceed on the understanding that it is in the public, and global, interest to facilitate such development as far as possible. It should not be in the power of a successor to property to hinder such change. However, it could be seen as an infringement on the rights of the proprietor of commercial property to prevent them from determining, to a certain extent, the use of their land. Part of the problem lies in the accessibility to the content of lease agreements. Bell and Rennie's *Journal* article cited above concerns mainly long leases,[37] which must be registered.[38] As such it is no hardship for a potential purchaser (aware of the lease's existence) to discover its terms.

However, for shorter leases there is a problem regarding the publicity of their content. Ultimately, it is bound in an agreement between the original contracting parties, the details of which may be kept private. In addition, "there may be certain obligations undertaken by the original parties

[33] *Bissett v Magistrates of Aberdeen* (1898) 1 F 87.
[34] WS Society Property Law Panel, The Signet Library, Edinburgh, 26 October 2017.
[35] Unfortunately, at time of writing the SLC has no material available on this consultation.
[36] D Bell and R Rennie "Purchase Options in Long Leases" (2006) 51(5) Journal of the Law Society of Scotland 49.
[37] *Ibid.*
[38] Registration of Leases (Scotland) Act 1857 s 20C.

to a lease which are purely personal"[39] and therefore not normally binding on successors. Of course, it could be stipulated in the original contract of lease that all conditions are to be binding on successors, but whether that would hold sufficient weight with a court is questionable. Nor does it solve the issue of the purchasers' knowledge of content. To address the problem of transparency when transacting with property subject to a commercial lease, the twenty-first century offers a simple technological solution: an online register. It would not be difficult to establish such a register of commercial leases, searchable by interested parties. Nor would it be difficult to administer, if parties were able to upload leases themselves. Negotiation of commercial leases already takes time and consideration, especially if there are specific clauses the parties wish to include peculiar to their lease. It would not be too expansive an approach to require subsequent registration of the leases.

There are some reservations about the proposition to permit all terms of commercial leases to be binding on successors. It could have a serious impact on the commercial property market. If, say, options to purchase were made binding on a new proprietor, it makes the purchase less attractive as the new owner may be divested of their property at a future date. Therefore, the market price becomes depressed. Also, it could create a situation dangerously close to feudalism in the industry. So, to balance environmental needs, and to encourage the commercial property market, I propose that an online register should be established for commercial leases concerning energy production, and that only such leases should be binding in their entirety on singular successors.

E. CORPOREAL MOVEABLE PROPERTY: A STEP TOO FAR?

So far, we have considered situations in which parties should be able to create their own real rights in relation to heritable property. What then of corporeal moveable property? What if the Giant gave Jack the goose that laid the golden eggs, and the gift was absolute? We need not go too far into the reasons he may do this, but for arguments' sake, let us suppose that he did not wish for future responsibilities regarding the goose (veterinary bills, and so on), but nonetheless wished to retain an annuity of a single golden egg per annum from her yield. Furthermore, what if the Giant wished to retain this annuity regardless of whether Jack continued to own the goose, so long as the goose remained alive and productive?

This type of retentive right could be characterised as a conditional right to the fruits of the moveable object: the goose. If we were to follow the South African model, with land this would be permissible as the obligation would be on the property,[40] and not personal in nature. However, the goose is not heritable. It is not registrable. It would be near impossible for any successor parties to identify whether or not such conditions were attached to its ownership. Unlike commercial leases, it would be a hindrance to attempt to establish an online register for transactions of moveable property, as such transactions normally are executed in person, without legal counsel, and are far too numerous to render registration practical. If one was unable to purchase a newspaper without registration, for example, circulation of print media would be irretrievably damaged.

Furthermore, it is difficult to see where the advantage of permitting such real rights to be created would lie. A goose that lays golden eggs is a thing of fantasy. Perhaps livestock would offer a similar situation in which such a condition might be applicable. A farmer may wish to retain a right to one calf in every four produced by a cow he sells. The benefit of this, however, is limited. A cow produces on average only seven to ten calves in a lifetime,[41] and the additional costs associated with trying to ascertain whether such a right affected cattle would outweigh the slim advantage to the seller. As such, although in certain situations with heritable property parties should be able to create

[39] Bell and Rennie (n 36).
[40] *Pearly Beach Trust v Registrar of Deeds* 1990 (4) SA 614 (C).
[41] F Murray, "Record Breaker? Mother of all cows gives birth to 16 calves", available at *http://www.bbc.co.uk/news/world-europe-21655215.*

their own real rights, with corporeal moveable property it is impractical, burdensome, and a step too far.

F. CONCLUSION

The idea of having a limited list of available real rights stems from the idea that parties transacting with property should have certainty, clarity, and freedom. It is asserted that to permit parties to create their own real rights would bring additional costs, which in turn would be detrimental to the market economy. However, as I have shown, this has not proven to be the case in South Africa, nor should it be the case in Scotland should the proper safeguards be put in place. Concerns that untrammeled expansion of rights could be the result of the relaxing of the rules regarding the *numerus clausus* principle are overblown.

Scotland may claim to have a fixed list of available real rights, but the reality is that if the list may be added to by discovery by the courts or by legislation, this cannot honestly be said to be true. We should not continue to claim to tether ourselves strictly to a Roman Law model when considering real rights, especially when 2000-year-old doctrines may not be wholly appropriate in the modern world. Nor should we leave it to the judiciary or legislature to be the creators of such rights, as they can be slow, reactive, and out of touch with commercial needs.

In South Africa, strict conditions on which real rights may been created have been effective and have permitted expansion only in circumstances where it has been necessary. These circumstances have not been numerous. Register-ability has been key to this. As such, in Scotland we should allow parties to create real rights in relation to heritable property. These real rights should be extended to personal servitudes, provided that the property and parties are adequately identified in the Land Register. Furthermore, there should be a register of commercial leases established for leases relating to energy production, and these leases should be binding in their entirety on singular successors. However, creation of real rights should not be extended to corporeal moveable property, as this would be too cumbersome to allow the free flow of the market economy. It would be, essentially, to grow a beanstalk too tall to climb.

DIRECT DISCRIMINATION: REDEMPTION BY GENERAL DEFENCE?

*Rebecca C E Gray**

A. INTRODUCTION

Currently, UK equality legislation separates direct discrimination from indirect discrimination in that the former cannot be justified by a general defence. By definition,[1] indirect and direct discrimination are distinct legal concepts with different intentions; direct discrimination seeks to achieve formal equality while indirect discrimination aims to find substantive equality.[2] Lady Hale explains this particularly well:

> The basic difference between direct and indirect discrimination is plain... The rule against direct discrimination aims to achieve formal equality of treatment: there must be no less favourable treatment between otherwise similarly situated people on grounds of colour, race, nationality or ethnic or national origins. Indirect discrimination looks beyond formal equality towards a more substantive equality of results: criteria which appear neutral on their face may have a disproportionately adverse impact upon people of a particular colour, race, nationality or ethnic or national origins.[3]

In a 2014 speech given by Lady Hale at the Comparative and Administrative Law Conference at Yale Law School,[4] she discusses the "clash of equality rights";[5] in particular, the relationship between the protected characteristics of sexual orientation and religion or belief.[6] She proposes extending the general defence of justification to cases of direct discrimination. Lady Hale posits that this would empower the courts and tribunals to look past the *prima facie* facts of direct discrimination

*4th Year LLB student, University of Edinburgh.
[1] Direct discrimination is defined in Equality Act 2010 s 13; indirect discrimination is defined in Equality Act 2010 s 19.
[2] H Collins, "Discrimination, Equality and Social Inclusion" (2003) 66 MLR 16 at 16-17.
[3] *R (E) v Governing Body of JFS* [2009] UKSC 15, [2010] 2 AC 728, 757B-C at para 56 per Baroness Hale.
[4] Lady Hale, "Religion and Sexual Orientation: The clash of equality rights" (2014), available at *https://www.supremecourt.uk/docs/speech-140307.pdf.*
[5] *Ibid.* at 1.
[6] Equality Act 2010 s 4.

cases and address "the real issues – legitimate aim, rational connection, [and] proportionality".[7] Furthermore, the differences between the anti-discrimination regimes under EU law and the ECHR are discussed. Lady Hale suggests that the ECHR system, which carries an open list of protected characteristics[8] and the ability to defend all discrimination actions, is preferable. Indeed, Lady Hale notes that she has been "arguing for a long time now that the ECHR approach is preferable to the EU approach".[9] This paper will examine this claim with a view to determining its validity.

(1) Brief overview of the development of UK discrimination law

The UK Equality Law framework has been developing rapidly since the introduction of a plethora of new Acts in the 1970s.[10] The pace of change has been relatively rapid but fragmented, with a complex web of legislation, statutory instruments and statutory codes of practice in force before the introduction of the Equality Act 2010. This new Act sought to consolidate and simplify the increasingly inaccessible legislation, and was drafted with the intention of being accessible to the workers it seeks to protect. Whilst much of the UK equality legislation has been more progressive and further reaching than required by the EU, certain elements have come directly from the European Directives.[11] The lack of a general defence for direct discrimination is one such element.[12] Though there are jurisdictions beyond the borders of the EU with the same absence of a general justification defence to direct discrimination,[13] there are also various jurisdictions which do allow general defences to direct discrimination. These countries include the United States of America and Canada,[14] whose defences go beyond the limited few available to EU member states.[15] Both of these jurisdictions follow the equality regime of the ECHR and, as such, allow citizens to justify and defend direct discrimination.[16]

(2) Issues with direct discrimination

There are four different kinds of prohibited conduct in the UK equality law framework designed to disincentivise and penalise discriminatory behaviour: direct discrimination, indirect discrimination, harassment/sexual harassment and victimisation. Protection against direct discrimination is designed to prevent overt discriminatory conduct by employers and so defences for employers are limited and set out in narrow statutory terms. In contrast, indirect discrimination, which seeks to resolve

[7] Hale (n 4) at 6.
[8] Article 14 has an "other status" which allows new characteristics to be added. This contrasts with the closed list of protected characteristics under the EU regime as in Council Directive 2000/78/EC OJ 2000 L303, and in Equality Act 2010 s 4.
[9] Hale (n 4) at 5.
[10] For example, the Equal Pay Act 1970; Race Relations Act 1976; and Sex Discrimination Act 1975, amongst other statutes and statutory instruments.
[11] One such element being the closed list of nine protected characteristics which come from art 19 of the Treaty on the Functioning of the European Union, Directive 2000/43/EC OJ 2000 L180, Directive 2000/78/EC OJ 2000 L303, Directive 2006/54/EC OJ 2006 L204.
[12] As explained in European Union Agency for Fundamental Rights, "Handbook on European non-discrimination law" (2010) at 21, available at *https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=3&ved=0ahUKEwiJotndmp7bAhXHK8AKHcg3DiwQFgg2MAI&url=http%3A%2F%2Ffra.europa.eu%2Fsites%2Fdefault%2Ffiles%2Ffra_uploads%2F1510-FRA-CASE-LAW-HANDBOOK_EN.pdf&usg=AOvVaw18CWfjES18zz16vUr5425v*.
[13] Australia is an example of a jurisdiction with no general defence to direct discrimination, merely limited exemptions and exceptions, Australian Human Rights Commission, "Direct Discrimination", available at *https://www.humanrights.gov.au/quick-guide/12026.*
[14] For example, in Canada, there is a duty to make reasonable accommodation based on any of the 13 grounds identified in Canadian Human Rights Act (CHRA) s 2, but with no obligation for people to act in a way contrary to their religious beliefs. From the Government of Canada, "Duty to Accommodate: A General Process For Managers", available at *https://www.canada.ca/en/treasury-board-secretariat/services/values-ethics/diversity-equity/duty-accommodate-general-process-managers.html*.
[15] Such as the Schedule 9 occupational exemptions in the Equality Act 2010.
[16] As explained in European Union Agency for Fundamental Rights, "Handbook on European non-discrimination law", 2010 at 21.

discriminatory provisions, criteria and practices, has a general defence. This defence provides the employer with the opportunity to justify their behaviour on the grounds of proportionality in achieving a legitimate aim.[17] Whilst the difference in defence tends to be attributed to the separate aims of direct and indirect discrimination, there is no uniform agreement about this among members of the judiciary and legal scholars. There are schools of thought which rigidly adhere to the premise that allowing justification to discrimination based on protected characteristics is beneath "human dignity",[18] while others view the current statutory position as untenable due to the frequent conflict of protected characteristics.[19] It seems that however persuasive the policy reasons for having no general defence, there are three strong arguments which clearly demonstrate the critical necessity of reform of direct discrimination: (1) the issue of competing protected characteristics, (2) the problem of overlap between direct and indirect discrimination, and (3) the need for continual revision of the protected groups.

On the other hand, persuasive statements illustrating the current inadequacies do not necessarily equate to a call for the introduction of the defence. An alternative remedy might be found through the open-list approach of the ECHR, or by the implementation of a reasonable accommodation defence.[20] However, it is submitted that the introduction of a general defence is crucial, in order to justify the continuation of the existing judicial mechanisms for discrimination in employment.

B. FUNDAMENTAL PROBLEMS WITH THE LACK OF A GENERAL DEFENCE

(1) Response to the human dignity argument

Many scholars equate a justification to direct discrimination to an attack on the fundamental principles of equality law. Indeed, Hepple describes a justification defence as "a dangerous heresy […] threatening to subvert the developing principle of equality in Community law".[21] However, there are so many narrow exceptions and exemptions to discrimination of the protected characteristics that it seems short-sighted to disregard a general defence that may have the ability to simplify the law.[22] If human dignity was so attached to formal equality, surely there would be much more advocacy for no defences or exemptions? For example, under the Schedule 9 exemptions in the Equality Act 2010, it would be possible to be exempt from the Act's requirements for the purposes of maintaining the continuity of a character in a theatrical production.[23] If the aim of this Act was indeed human dignity, it leads one to question why such an exemption is available for somewhat trivial purposes. A general defence would not necessarily allow discrimination beyond the circumstances currently permissible but would simply offer the judiciary greater flexibility in examining the facts of each case and in exercising their own judgment rather than being bound by legislation. Thus, it seems difficult to argue that a general defence of direct discrimination risks violating human dignity, while simultaneously being ideologically at ease with the narrower defences currently available.

---

[17] Indirect discrimination can be defended if the employer can show the action was a proportionate means of achieving a legitimate aim, see Equality Act 2010 s 19(2)(d).
[18] T Gill and K Monaghan, "Justification in Direct Sex Discrimination Law: Taboo Upheld" (2003) 32(2) Industrial Law Journal 115.
[19] Hale (n 4) at 20 "I find it hard to believe that the hard-line EU law approach to direct discrimination can be sustainable…".
[20] Government of Canada, "Duty to Accommodate: A General Process For Managers", available at *https://www.canada.ca/en/treasury-board-secretariat/services/values-ethics/diversity-equity/duty-accommodate-general-process-managers.html*.
[21] B Hepple, "Can direct discrimination be justified?" (1994) 55 Equal Opportunities Law Review 48.
[22] J Bowers QC and E Moran, "Justification in Direct Sex Discrimination Law: Breaking the Taboo" (2002) 31(4) Industrial Law Journal 307.
[23] Schedule 9 Occupational Exemptions would allow discrimination by casting a male actor ahead of an actor of another gender.

The human dignity argument is weak, as it relies on the assertion that discrimination damages human dignity.[24] However, there are many examples of discrimination that are completely innocuous and do nothing to undermine human dignity, such as religious discrimination for religious occupations.[25] Bowers and Moran consider the utility of a general justification defence in light of the then-recent introduction of a general defence to discrimination against employees on fixed-term contracts.[26] From this perspective, they considered why there was reluctance to a general defence when there were already numerous narrow defences. This is demonstrated in the exceptions to occupational requirements,[27] which specify that the Equality Act 2010 will have no effect under certain circumstances. Here, the implication is that an employer can insist that a worker possess certain qualities, even to the detriment of the protected characteristics.[28] For example, there would be an exception for the purposes of safeguarding national security.[29] The employer has to show that the requirement is occupational, that the application of the requirement is a proportionate means of achieving a legitimate aim, and that the person to whom the requirement is applied cannot meet it (or that the employer has reasonable grounds for not being satisfied that the person meets it).[30] This idea of innocent motivations in discrimination is discussed further by Dworkin, who foresaw its necessity in certain cases.[31]

Further, the notions of positive action, positive discrimination and affirmative action merit discussion. These are all ways in which protected groups might be treated more favourably in order to boost the likelihood of equality of outcome i.e. substantive equality.[32] This tends to be employed to limit indirect discrimination. However, treating a protected group more favourably necessarily means less favourable treatment of those in different protected groups or those without any relevant characteristics, i.e. guaranteeing interviews to disabled applicants means that all non-disabled applicants have a higher threshold to meet. In this scenario, non-disabled people are not a protected group. A further example arises when a company guarantees interviews to those with a minority racial background, which would discriminate against the majority because of their race.[33] This is problematic for the human dignity argument; if characteristics must be protected at all costs, how can substantive equality be sought without interfering with the dignity of certain groups? This is an important consideration as positive discrimination measures are central to redressing the imbalance in the labour market caused by structural disadvantage and long-term systematic discrimination. Positive action is essential in providing redress to protected groups who have long been disadvantaged in the labour market. Therefore, the human dignity argument is counter-productive against the substantive equality necessary to build an equal-opportunities work environment. Whilst human dignity might be a factor that promotes equality law, it should not be used as a definitive reason against the introduction of a general justification defence.

---

[24] As argued by Gill & Monaghan (n 18) at 115.
[25] Equality Act 2010 Schedule 9 paras 7-16.
[26] Bowers QC & Moran (n 22) at 307; Fixed-Term Employees (Prevention of Less Favourable Treatment) Regulations 2002 s 3(3)(b).
[27] Now found in the Equality Act 2010 Schedule 9, then found in Article 4(1) of the Framework Directive, Article 4 of the Race Directive, and Article 14(2) of the Recast Equality Directive.
[28] Protected characteristics of sex, age, sexual orientation, marriage and civil partnership and religion and belief can all be discriminated against under the Schedule 9 Work Exceptions in the Equality Act 2010.
[29] D Cabrelli, *Employment Law in Context: Text and materials,* 2nd edn (2016) 455.
[30] Para 1(1) of Schedule 9 of the Equality Act 2010 provides that, in the context of work certain characteristics can be lawfully discriminated against.
[31] R Dworkin, "Reverse Discrimination", in Duckworth (eds), *Taking Rights Seriously* (1977) ch 9, discusses this in the context of race discrimination.
[32] And thus, there will be discrimination against 'certain mainstream groups' per Cabrelli, *Employment Law* 450.
[33] There is no specific example of racial discrimination in the new occupational requirements of Schedule 9, but this perhaps raises the question of whether there ought to be, as this would be a good way of tackling indirect discrimination.

(2) Overlap between direct and indirect discrimination

The frequent overlap in case law between direct and indirect discrimination presents an additional problem. This is due to the difficulty of establishing whether the discrimination can be attributed to a protected characteristic over an unprotected characteristic, i.e. the difficulty of ascertaining if a provision, criterion, or practice (PCP) is neutral or not. This is why, in part, the courts have been using a "reason why" test.[34] Both direct and indirect discrimination involve protected groups being treated differently due to a particular characteristic. However, in direct discrimination the treatment stems from a PCP using a specific protected characteristic as a measure,[35] i.e. bonuses only being awarded to men. This is distinct from indirect discrimination, where the equivalent would be bonuses which are only awarded to full-time workers. This might constitute indirect discrimination because part-time workers tend to be disproportionately women.[36] In cases of indirect discrimination, the PCP is unrelated to a protected characteristic, but results in a disproportionate disadvantage on those belonging to a group with a certain characteristic. There are numerous cases sitting between the two types of discrimination due to the difficulty in maintaining "the hard and fast demarcation between direct and indirect discrimination".[37] The cases below are particularly strong illustrations of this difficulty in ascertaining direct or indirect discrimination.

In *James v Eastleigh,*[38] a swimming pool discounted their price on the basis of pensionable age. This had been equalised in 1989 (after the event), and thus the court found that, whilst pensionable age was not a protected characteristic, it was so closely intertwined with sex that they were able to apply the "but for" test from *Birmingham City Council v EOC*[39] and held that sex discrimination had taken place. In Lord Bridge of Harwich's judgment, he found the following:

> "Lord Goff's test, it will be observed, is not subjective, but objective. Adopting it here the question becomes: 'Would the plaintiff, a man of 61, have received the same treatment as his wife but for his sex?' An affirmative answer is inescapable".[40]

There was clearly no intention to discriminate on grounds of sex, and whilst intention is not a defence,[41] this case is controversial and arguably lies outwith the intention of the Act. Indeed, direct discrimination was found to be present in this case, though Lord Griffith and Lord Lowry dissented. Lord Lowry writes a lengthy opinion in which he proposes that there must be some subjectivity to the test due to the necessity of subjectivity in discrimination. He further adds that the case could not be indirect discrimination under the statutory regime then in force;[42] however, had the case fallen under the Equality Act 2010 there may have been a very different result. The case seems to be a textbook example of indirect discrimination with a neutral PCP disproportionately affecting a protected group in the current regime of equality law.

---

[34] In direct discrimination cases, the court looks beyond the facts *prima facie* and seeks to establish if it was the protected characteristic that was being targeted as opposed to another unprotected characteristic. This developed in the case of *Nagarajan v London Regional Transport* [2000] 1 AC 501 and *Shamoon v RUC* [2003] ICR 337, from the earlier 'but for' test in *Birmingham City Council v Equal Opportunities Commission* [1989] AC 1155 at 1193-1194 per Lord Goff of Chieveley, and developed in *James v Eastleigh* [1990] 2 AC 751.

[35] *R (E) v Governing Body of JFS* [2009] UKSC 15, [2010] 2 AC 728, 757B-C at para 56 per Baroness Hale.

[36] 41% of women work part-time in comparison to 11% of men, Office for National Statistics, "Annual Survey of Hours and Earnings 2015 Provisional Results" (2015), available at *https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/bulletins/annualsurveyofhoursandearnings/2015provisionalresults*.

[37] *Strathclyde Regional Council v Wallace* [1998] ICR 205 at para 213b per Lord Browne-Wilkinson.

[38] *James v Eastleigh* [1990] 2 AC 751.

[39] *Birmingham City Council v Equal Opportunities Commission* [1989] AC 1155.

[40] *James v Eastleigh* [1990] 2 AC 751 at para 765 per Lord Bridge of Harwich.

[41] "The intention or motive of the defendant to discriminate…is not a necessary condition of liability," *Birmingham City Council v Equal Opportunities Commission* [1989] AC 1155 at paras 1194A-1194D per Lord Goff; there is also, of course, no defence to direct discrimination.

[42] It did not meet the requirements under s 1(1)(b) Sex Discrimination Act 1975, which was then in force.

The case of *Patmalniece v Secretary of State for Work and Pensions*[43] concerns a "right to reside" test which had to be satisfied by certain welfare benefit claimants. All UK citizens were able to satisfy this test, but citizens of other EU countries had to prove additional qualifications. Lady Hale proposes that this is an example of an incident where it is hard to discern whether there is direct or indirect discrimination.[44] Mrs Patmalniece argued that there was both direct and indirect discrimination due to the direct discrimination on grounds of nationality in the "right to reside" test, in addition to indirect discrimination in the related legislation. A discussion was had about what constitutes direct discrimination, with reference to the commentary in *Bressol v Gouvernement de la Communauté Française.*[45] Lord Walker suggested that the best guidance comes from the opinion of Advocate General Jacobs,[46] who stated that "the discrimination is direct where the difference in treatment is based on a criterion which is either explicitly that of sex or necessarily linked to a characteristic indissociable from sex".[47] The case demonstrates the practical difficulty of determining a distinction between direct and indirect discrimination. All judges except Lord Brown delivered lengthy judgements, each weighing in on how the definition of direct discrimination was assessed in other European case law. The residency test was held justifiable as an indirectly discriminatory measure. The commentary in this case demonstrates a situation aptly described by Lady Hale, whereby such exhaustive discussions often mean that the courts lack the time to address "the real issues".[48] Had a defence to direct discrimination been available, the time spent and jurisprudence generated would have been avoided, and the court may have been able to comment on "the real issue" concerning the legitimacy of limiting state benefits.[49] In a post appearing on the Supreme Court's blog, Christopher Brown makes an important proposition: "perhaps it is time for the ECJ's insistence that direct discrimination on nationality grounds can never be justified to be revisited".[50]

The two recent cases of *Achbita*[51] and *Bougnaoui*[52] concerned female Muslim employees unable to wear headscarves at work without risking dismissal. In *Achbita*, there was a prohibition on employees wearing symbols of political, philosophical or religious significance in the workplace. In *Bougnaoui*, there was no internal rule, but the employer insisted upon the removal of the headscarf due to apparent customer upset. In both cases, there was discussion about the neutrality of the PCPs, with particular confusion in *Bougnaoui*. The divergent opinions of Advocates General Kokott and Sharpston serve as further evidence of the difficulty in establishing clear judicial reasoning in borderline cases.[53] In *Achbita*, it was submitted that that the national court had misconstrued the law by finding no direct discrimination. This was due to the PCP applying to political, philosophical and religious belief symbols as opposed to just religious symbols or even just symbols of Islam. However, the CJEU declined to respond on the basis that it was a matter for the national courts to decide. There is obviously great discrepancy and confusion in this area, with courts in France and the UK finding there was indirect discrimination, Belgium finding direct discrimination and G4S (the employer in *Achbita*) submitting no discrimination. This uncertainty in the law calls for an urgent need for clarification by the CJEU on the rigid steps which must be taken to distinguish direct and indirect discrimination, despite their clear reservation in doing so.

---

[43] *Patmalniece v Secretary of State for Work and Pensions* [2011] UKSC 11.
[44] Hale (n 4) at 6.
[45] Case C-73/08 *Bressol v Gouvernement de la Communauté Française* [2010] 3 CMLR 559.
[46] *Patmalniece v Secretary of State for Work and Pensions* [2011] UKSC 11 at para 66.
[47] Case C-79/99 *Schnorbus v Land Hessen [2000] ECR I-10997* at para 33 per Advocate General Jacobs.
[48] Hale (n 4) at 6.
[49] Although this was addressed but not at a length which equated to that taken by discussion of direct and indirect discrimination.
[50] C Brown, "Case Comment: Patmalniece v SSWP [2011] UKSC 11", available at http://ukscblog.com/case-comment-patmalneice-v-sswp-2011-uksc11/.
[51] Case C-157/15 *Achbita & Anor v G4S Secure Solutions NV* [2017].
[52] Case C-188/15 *Bougnaoui and ADDH v Micropole SA* [2015].
[53] Case C-157/15 *Achbita & Anor v G4S Secure Solutions NV* [2017] per Kokott; *Ibid*. per Sharpston.

There is considerable narrowness in the distinction between direct and indirect discrimination in many of the cases which make it to court. It seems therefore, that regardless of the pronounced ideologies which initially lead to different defence positions, in practice there is often more similarity than difference in direct/indirect discrimination cases. The available defences ought to reflect this similarity. Further, because the difference is so difficult to distinguish and there is no general justification defence for direct discrimination, there is a strong incentive to bring cases to court under direct discrimination claims. This is due to the difficulty the employer will have when attempting to justifying any action based on direct discrimination. Had *James v Eastleigh* been brought as an indirect discrimination case, there is little doubt the Council would have been able to justify its actions as a form of positive discrimination.

(3) Necessity of continual revision of protected characteristics

The Equality Act 2010 protects the nine characteristics of age, disability, gender reassignment, marriage and civil partnership, pregnancy and maternity, race, religion or belief, sex and sexual orientation.[54] This list codifies characteristics that have been found to be worthy of protection at different points between the introduction of direct discrimination law in the 1970s and the drafting of the Equality Act (consultancy in respect of the latter had begun in 2005).[55] However, the current closed list means that the judiciary have no discretion to assess alternative characteristics and protect them. Instead, either they deny protection to vulnerable parties or they associate an unprotected characteristic with one that is, thereby ensuring protection. This contributes to the growing need for serious review and reform of discrimination law and proves that there are problems with the current regime beyond the need for a defence advocated by Lady Hale.

The case of *Onu v Akwiwu*[56] is a strong illustration of the difficulty that vulnerable parties have of finding protection when the protected characteristics are rigidly fixed. The case concerned a female who claimed that she was subject to abuse and discrimination by her employer due to her migrant worker status. The court found there could be no direct discrimination as this was not a protected characteristic. There was much discussion about other reasons for the abuse and there was difficulty in relating the migrant worker status to the protected characteristics. It was said there were at least twenty cases waiting to be heard on the basis of the ruling. There was also the case of *Taiwo v Olaigbe,*[57] which involved vulnerable migrant workers who were treated badly by their employers.[58] Together these cases were examined by the Supreme Court,[59] which found that the treatment was not directly discriminatory because, whilst the mistreatment was due to their migrant status, it could not be equated with race or nationality.[60] Further, there could be no indirect discrimination because there was no exercise by the employers of a discriminatory policy, criterion or practice relating closely enough to a protected characteristic. There is clearly a problem when the court can recognise that there is a characteristic which makes workers vulnerable and prone to frequent exploitation by their employers but can offer no remedy in discrimination law.

It is arguable that the position in *Onu v Akwiwu* is similar to that of *James v Eastleigh*, although in the latter the court was able to equate the pensionable age to sex. It appears that, had there not been so much judicial development in the intervening period between these two cases, a different conclusion may have been reached in the migrant worker cases. If we look at the ECHR structure, factors such as socio-economic class and home address have been found to qualify as protected characteristics under the "other" category.[61] This is a weakness of the EU law approach, which is

[54] Equality Act 2010 s 4.
[55] Explanatory Notes to the Bill for the Equality Act 2010, para 7.
[56] *Onu v Akwiwu* [2014] 1 WLR 3636.
[57] *Taiwo v Olaigbe and another* UKEAT/254/12, UKEAT/285/12.
[58] Many of whom claimed to be family members to try and shirk paying their workers the minimum wage.
[59] *Taiwo v Olaigbe* [2016] UKSC 31.
[60] *Ibid.* at para 11 per Lady Hale.
[61] Art 14 of the ECHR.

inferior to that of the ECHR advocated by Lady Hale.[62] It is an element of UK equality law that ought to be reformed alongside the introduction of a general defence for direct discrimination cases.

(4) Issue of competing protected characteristics

The clashing of two protected characteristics is arguably the most pressing reason why the introduction of a general justification defence to direct discrimination is necessary. There is an alarming regularity with which cases of apparent discrimination on the grounds of sexual orientation by Christians are reported in the media. One commonly reported situation is that of a Christian couple operating a guest house and refusing rooms to unmarried couples (as homosexual couples could not be married until recently, it was impossible to not discriminate against them).[63] There tends to be a finding of discrimination on the part of the Christian managers, making them the perpetrators of direct discrimination with no defence to their behaviour. That said, there is a very strong argument that the courts conduct discrimination against religious groups by denying their right to freedom of religion. Lady Hale can sense the inadequacy of the current regime and sees the need for a provision in respect of religious beliefs.[64] Here, Lady Hale compares Canada to England, noting that, "a country with an established church [is] less respectful of religious feelings than one without".[65] A general justification defence would allow religious believers the opportunity to faithfully follow their beliefs and use an alternative criterion to ensure a reasonable balance is struck between clashing protected groups. In the words of Lady Hale: "would it not be a great deal simpler if we required the providers of employment, goods and services to make reasonable accommodation for the religious beliefs of others"?[66]

A review of some recent UK cases regarding the clashing of two protected characteristics demonstrates the utility of a reasonable accommodation defence, either exclusively for cases where religious beliefs have clashed with other protected characteristics or more generally for all clashes of protected characteristics. The case of *Bull v Hall*[67] concerned the letting of a hotel room. The defendants let double rooms exclusively to married couples and twin and single rooms to anyone else. It was claimed that this was direct discrimination against homosexual couples. The defendants argued that there was no discrimination on the grounds of sexual orientation as they did not let to *anyone* who was unmarried. If any kind of discrimination was present, it was indirect discrimination and it could be justified by their right to freedom of religion. Following the reasoning in *James v Eastleigh*, at a time when homosexual couples could not marry, it was direct discrimination because the characteristic correlated exactly with the PCP, even though the PCP was *prima facie* neutral. This finding was later questioned in the case of *Black and Morgan v Wilkinson.*[68] The Master of the Rolls did not recognise that there had been direct discrimination, despite the finding in *Bull*. He advanced the view that this was indirect discrimination but found himself bound by the earlier judgment. The difficulties experienced in cases involving discrimination by religious groups arises from the protected characteristic of religion. It is difficult to protect both religion and the other party seeking equality. The issue of indirect and direct discrimination discussed earlier will clearly arise again as courts seek to offer fairness to both parties. However, the nature of the competing characteristics demands additional contemplation.

Arguably there is considerable damage being done through the stubbornness of policy makers to reform the prohibited conduct of direct discrimination. Religious believers, frequently of the Judeo-Christian tradition, are often told that their beliefs are indefensible when asserted against other protected groups (the latter protection usually being sexual orientation). This marginalisation of

[62] Hale (n 4) at 5.
[63] Same-sex marriage legislation took effect in December 2014 in Scotland and in March 2014 in England and Wales. Northern Ireland still does not permit same-sex marriage.
[64] Hale (n 4) at 20.
[65] *Ibid*.
[66] Hale (n 4) at 18.
[67] *Bull v Hall* [2013] UKSC 73.
[68] *Black and Morgan v Wilkinson* [2013] EWCA Civ 820.

religion could be eradicated with the introduction of either a general justification defence or a similar defence restricted to cases involving two parties with protected characteristics. This would ensure equal treatment and proportionality for all protected groups. Such defences would allow the courts to look more critically into mutual respect and fair treatment without trying to split hairs over indirect and direct discrimination according to whether they believe the appellants merit a defence. As in ECHR cases, there needs to be an attempt to seek balance rather than offer little or no protection to protected groups who discriminate.

C. **PROPOSED RESOLUTIONS**

(1) Does the ECHR approach better combat the four issues under discussion?

The ECHR equality regime is advantageous in many ways, particularly the balancing exercise that it employs between protected characteristics. An examination of *Eweida and Others v United Kingdom*[69] is a testament to that. This was a referral of four cases involving Christians seeking affirmation that their right to practice their religion under Article 9 had been limited without justification, or that they as individuals had been discriminated against contrary to Article 14 of the ECHR. Ms Eweida worked for British Airways and Ms Chaplin worked for the NHS. Both stated that they were not permitted to wear a cross. Ms Ladele worked for the London Borough of Islington and Mr McFarlane worked for Relate (a charity). They also complained that they were dismissed due to their views regarding homosexual relationships. Ultimately, only Ms Eweida had her complaint upheld by Strasbourg. The essential aspect of the case is embodied in the assertion that: "regard must be had in particular to the fair balance that has to be struck between the competing interests of the individual and of the community as a whole".[70] Regardless of the outcome of these cases, the claimants' opportunity to have their right to practice their religion balanced against other community priorities, in addition to the consideration of proportionality under the Convention, was the most just and fair treatment they could have encountered under the various judicial channels available to them. It is, however, unjust to require the parties to appeal to the European Court of Human Rights to find this balanced approach.

A further advantage of the ECHR regime, beyond the balancing approach to competing protected characteristics, is the open list. Under Article 14, new vulnerabilities can qualify for protection if they fall under "other status". This offers the best chance of ensuring discrimination law continues to protect worthy vulnerable groups. This is a solution that ought to be considered in the UK due to its better ability to ensure all groups are given equal opportunity to defend and justify policies.

(2) The utility of reforming defences

In the Canadian system, when there is a clash of characteristics, there is a defence available if the would-be defendant can demonstrate that they have met their duty of reasonable accommodation. Lady Hale advocates this in her speech. She states, "would it not be a great deal simpler if we required the providers of employment, goods and services to make reasonable accommodation for the religious beliefs of others"?[71] This would involve an overhaul of the existing system and a move towards a point of balancing and defining what reasonable accommodation might look like in cases such as *Bull v Hall*. According to cases heard by the British Columbia Human Rights Tribunal, factors to be considered might include the effects of the action(s) on the holder of the opposing characteristic, apologising for inconveniences, reimbursement of any expenses lost, the offering of help to find a

[69] *Eweida and Others v United Kingdom*, App nos. 48420/10, 59842/10, 51671/10, 36516/10 [2013] ECHR 37.
[70] *Ibid*. at para 84.
[71] Hale (n 4) at 18.

workable solution,[72] dealing with any cancellation respectfully and the exploration of alternatives.[73] Whilst this solution involves a balancing act which is often felt to be beyond the scope of the court and a matter of public policy, it offers the chance to give two parties with conflicting characteristics an opportunity to treat one another with mutual respect and understanding. This might develop into a more rigid list of expectations after some development of case law. Superficially, this gives flexibility to allow judicial discretion where proportionality is present. However, there would also be a danger that precedent might create rigid conditions for compliance. Thus, the introduction of a reasonable accommodation defence might go somewhat to ensure a balance is struck between protected groups, whereas a general defence has much more promise as a solution to the current problems with direct discrimination.

(3) Is there something in between direct and indirect discrimination?

The frequent difficulty encountered by the courts of organising case law into clearly defined categories of either direct or indirect discrimination was discussed above. One solution to this problem is to introduce a defence, thereby curtailing borderline direct discrimination claims. However, it would also be possible to create a new category of discrimination aimed at those cases that fall in between. This is discussed at length by Forshaw and Pilgerstorfer,[74] who propose a quasi-direct discrimination category following the *Enderby* case.[75] Here, the courts recognised that there was indirect discrimination taking place; however, the authors postulate that it may have been a form of direct discrimination given discussion around comparators and the fact that there was no neutral PCP which resulted in the inequality. The presence of both direct and indirect discrimination qualities but no ringing bell which firmly defines the discrimination as one or the other, suggests that the law has inadequacies. In each respective case discussed by the authors, it was noted that there lacked an ability for the claimant to use a separating rule to show a group "winning" to the claimant's "losing". The cases in question are narrow and tend to relate to equal pay cases.[76] This illustrates the tendency of the courts to shoehorn cases into a single category where claims arise with similar facts, rather than critically distinguishing the cases on their particular facts to establish if there is direct or indirect discrimination present. Whilst there is merit in the intellectual process employed by Forshaw and Pilgerstorfer, and their conclusions are measured, the idea of further complicating the law of discrimination is unattractive. It will not solve all of the issues delineated by Lady Hale,[77] such as the competition between protected characteristics. Thus, the category of quasi-direct discrimination is a useful solution but alone does not present an adequate one. It does not possess a strong ability to offer such far-reaching reform and improvement as a general justification defence.

D. CONCLUSION

There are numerous issues with the current approach to direct discrimination. Ultimately, the court frequently struggles to find a separation between the concepts of direct and indirect discrimination. This disparity presents a problem for defendants, who will either be able to justify their action under indirect discrimination or be near-defenceless under direct discrimination. This has the potential to skew cases coming to the court, with claimants favouring direct discrimination actions due to its perceived advantage for their claim. Further, the EU closed list limits the protection the court can

[72] The factors listed prior to the citation are from *Smith and Chymyshyn v Knights of Columbus and Others* 2005 BCHRT 544.
[73] From *Eadie and Thomas v Riverbend Bed and Breakfast and others (No 2)* 2012 BCHRT 247.
[74] S Forshaw and M Pilgerstorfer, "Direct and indirect Discrimination: Is There Something in between?" (2008) 37(4) Industrial Law Journal 347.
[75] *Enderby v Frenchay Health Authority and another* [1994] 1 All ER 495.
[76] *Enderby v Frenchay Health Authority and another* [1994] 1 All ER 495*; Secretary of State for Trade and Industry v Rutherford and Bentley (no.2)* [2006] UKHL 19; *British Medical Association v Chaudhary* [2007] EWCA Civ 788; *Allen v GMB* [2008] IRLR 690.
[77] Hale (n 4).

offer workers through equality law and means that workers may have to bring costly human rights cases to find a resolution to certain discriminatory circumstances. An example could be where migrant workers are treated unfavourably. Moreover, the potential for cases of competing protected characteristics to arise leaves the court in a position where precedent fails to demonstrate adequate consideration of the characteristic possessed by the defendant against the appellant. Such issues clearly illustrate the shortfalls of an impaired equality regime, which can only naturally lead to reform.

Lady Hale posits that a general defence to direct discrimination would resolve many of the issues under the current statutory regime. Alternative solutions might offer some relief – a new quasi-direct discrimination status, or a reasonable accommodation defence. It is the opinion of the author that either a general defence or a reasonable accommodation defence may be appropriate; however, history has demonstrated that circumstances change continually and thus a general justification defence is ultimately the best fit for a reform that will stand the test of time. A general defence will allow equality in the pursuit of discrimination cases between those of direct and indirect discrimination, as well as ensuring that the court is given an opportunity to evaluate the strength of defences relating to protected characteristics. Whilst Lady Hale notes that she is not sure "how comfortable [she] would be with the sort of balancing exercise required by the Canadian approach [duty of reasonable accommodation]",[78] she also states "courts and tribunals have a natural eye for what they see as the merits of the case".[79] It is entirely possible and desirable to introduce a general defence of justification to the accusation of direct discrimination in equality law.

[78] Hale (n 4) at 20.
[79] Hale (n 4) at 5.

BEHIND THE LAW: A FEMINIST CRITIQUE OF CRIMINOLOGICAL ENQUIRY

*Rebecca Drummond**

A. INTRODUCTION

While seemingly immortalised in law, crime is not a necessary, metaphysical manifestation, despite large claims by criminologists that its study can be value-neutral. Indeed, criminology is not, nor will it ever be, an exact science, as no laws apply to human behaviour in the manner with which they can apply to nature. As Raymond Michalowski points out, the law operates as part of the "ideological apparatus of the political state by taking state definitions of what is crime and who are its criminals as the starting points for criminal enquiry".[1] In the past, there have been a number of approaches to criminology. Positivists have argued that the explanation for deviant behaviour lies in the biological enquiry of pathology.[2] Sociological approaches consider cultural and structural factors - such as exposure to extreme poverty, broken family units, or childhood abuse - pivotal to an individual's criminal involvement.[3] More recently, feminist criminologists have linked criminality to the concept of "doing gender" where "gender systematically corroborates that identification through social interaction".[4] All three approaches question the common theme of sex (male) and class (lower) as disproportionate indicators for whether an individual becomes involved in criminal activity. However, in all three approaches, little to no inquiry is made into whether the law, as it stands, is skewed to

*4th year politics student at The University of Edinburgh.

[1] R Michalowski, "What is crime?" (2016) 24 Critical Criminology 181 at 181.

[2] For example, see D Klein, "The Etiology of female crime" in J Muncie, E McLaughlin and M Langan (eds), *Criminological Perspectives: A Reader* (1996); C Lombroso and W Ferrero, "The criminal type in women and its atavistic origin" in J Muncie, E McLaughlin and M Langan (eds), *Criminological Perspectives: A Reader* (1996); J Wilson and R Herrnstein, *Crime and Human Nature* (1986).

[3] D Downes and P Rock, "Culture and subculture" in D Downes and P Rock (eds), *Understanding Deviance: a guide to the sociology of crime and rule-breaking,* 5th edn (2007); W Miller, "Lower class culture as generating a milieu of gang delinquency" (1958) 14 Journal of Social Issues 5.

[4] J Messerschmidt, "Schooling, masculinities and youth crime by white boys", in T Newburn and E Stanko (eds), *Just Boys Doing Business? Men, Masculinities and Crime* (1995) 172; R Collier, "Law, sex and masculinity" in R Collier (ed), *Masculinity, Law, and the Family* (1995) 87; R Collier, "Masculinities and crime" (1998) 34 Criminal Justice Matters 21; L Gelsthorpe and A Morris, "Feminist perspectives in criminology: Transforming and transgressing" (1991) 2 Women & Criminal Justice 3.

produce such a result. Should we question why males from lower socio-economic backgrounds seem inherently predisposed to criminal behaviour, or critically examine whether the law is biased towards the poor, towards men, or indeed, towards poor men? I argue the latter; that in order to understand why people commit crime, we must first examine what we consider criminal behaviour, beyond its institutional, criminal justice definition. Therefore, I propose a feminist critique of criminological enquiry. However, rather than examining the role played by gender in deviant behaviour, I examine the role of gender in the definition of deviance, crime and justice. I argue that consideration must be taken of the political, economic, and historical context, which when combined, quarantine what we consider devious behaviour from the norm. I will analyse crime as constructed not only by a hegemonic, capitalist consciousness, but by a hegemonic, *masculine*, capitalist consciousness. Patriarchy is a political ideology, justified by law and legitimised by criminology. I first briefly define what is meant by gendering and hegemonic masculinity, before arguing against the possibility of objective criminological enquiry in the context of the United Kingdom.

B. THE DICHOTOMY OF SEX AND GENDER

There are countless theoretical frameworks within feminism, from liberalism, to socialism, to postmodernism. However, the point upon which feminists agree is the dichotomy between "natural sex", and gender. The idea of "natural sex" is a controversial one as those who are born with a different inward sex than that biologically assigned can feel excluded. Indeed, some feminists argue that women face discrimination based on their biological sex as opposed to their assigned gender, which they argue is socially constructed. Therefore, some feminists, for example Germaine Greer, have gone so far as to say that trans-women are not "real" women.[5] This is an opinion which flies in the face of the gender essentialism feminists have resisted against, and one I wholeheartedly disagree with. Therefore, while I recognise the concept of "natural sex" as a contention, for the sake of brevity, I take it to mean the biological sex with which an individual associates themselves.[6] While sex and gender are used interchangeably in everyday life, feminists argue that, contrary to sex, gender is a socially constructed, fluid concept which categorises qualities, characteristics and behaviours as either masculine or feminine.[7] Values such as chivalry and honour, and characteristics such as aggression or discipline, are often considered masculine. Similarly, mothering, nurturing, virtue and purity, are found to be associated with femininity. However, while masculinity is typically characterised to a far greater degree, femininity is often simply that-which-is-not-masculine, or "other". While gender is a category, there is also a performance element, what feminists term "doing gender".[8] The masculine act of doing gender tends towards the frame of hegemonic masculinity. This is a:

> cultural ideal [...] that commands the most respect, and influences how all men construct their masculinities [...] The value of the concept of hegemonic masculinity lies in […] why certain elite men have remained in positions of power and wealth, and, furthermore, how masculinised ideals have remained privileged in public life.[9]

Through this, the dominant position of men is legitimised, while the subordination and oppression of women is justified.[10]

---

[5] C Wahlquist, "Germaine Greer tells Q&A her trans views were wrong, but then restates them" (2016) *The Guardian,* available at *https://www.theguardian.com/books/2016/apr/12/germaine-greer-tells-qa-her-trans-views-were-wrong-but-then-restates-them.*

[6] For further reading surrounding this debate, the winner of the 2017 Catherine Stimpson Prize for Outstanding Feminist Scholarship is highly insightful: C Awkward-Rich "Trans, feminism: Or, reading like a depressed transsexual" (2017) 42 Journal of Women in Culture and Society 819.

[7] C Duncanson, *Forces for Good? Military Masculinities and Peacebuilding in Afghanistan and Iraq* (2013) 61.

[8] L Gelsthorpe, "Feminism and criminology" in M Maguire, R Morgan and R Reiner (eds), *The Oxford Handbook of Criminology* (2002) 112.

[9] R Connell and J Messerschmidt, *Hegemonic Masculinity: Rethinking the Concept* (2005).

[10] *Ibid.*

C. THE ROLE OF CRIMINOLOGY

Orthodox criminological enquiry is based on the premise that the law represents the norms of society. However, radical positivists raise pertinent questions challenging the concept of society, and whether it can be said to have a homogenous set of interests and norms. Pertaining to British society specifically, it seems naive to think that the societal evolution undergone over the last forty years has been duly reflected by changes in the law. Therefore, if the law does not reflect the norms of society, whose norms does it reflect? William Chambliss and Robert Seidman argue that "[t]he members comprising a social system are differentiated according to the positions they occupy in it [...] its content arises because the person or persons who occupy the position fulfil a complex of obligations and a complex of rights, associated with this position".[11]

Michalowski argues that in a capitalist society, these positions are hierarchical class structures, which reproduce obligations and rights "consistent with the needs of capitalist accumulation". I find his argument highly compelling, and he draws extensively on the issue of racial dynamics within the United States. While his argument does not specifically reference gender dynamics, their inclusion would not prove contradictory. Michalowski argues that in class-divided societies "the interests of dominant classes play a substantial role in determining what is and what is not a crime", and thus "a criminology that studies only behaviours criminalized by the state [...] becomes part of the social apparatus for reproducing existing class arrangements and their consequent inequalities".[12] Thus, we can revise Michalowski's argument to illustrate the feminist perspective. In a gender-divided society, the interests of the dominant gender play a substantial role in determining what is and what is not a crime. By not doing so, criminology and criminologists who concentrate only on behaviour become an inherent part of the social apparatus, reproducing existing gender arrangements and consequent inequalities. Patriarchy is the ideology, justified by law and legitimised by criminology.

D. THE GENDERED CONTEXT

In considering the political, economic and historical context of how we define criminality, we expose the impartial nature of the law and the resultant bias of orthodox criminology. Most notably, William Chambliss and Robert Seidman penned the critique *Law Order and Power* in 1971, arguing that "in all societies [...] those [groups] which are most likely to be effective are the ones that control the economic and political institutions [...] As a consequence, legislation typically favours the wealthier, the more politically active groups in the society".[13] Michalowski convincingly applies this theory to the class system, as did Taylor, Walton and Young who "attempted to eschew the scientific orthodoxy [of criminology] and instead illustrated how crime was socially constructed by the capacity of state institutions to define and confer criminality on others".[14] However, these attempts have been largely "ungendered", thus denying the significance of crimes committed against women as a pervasive element of oppression. Therefore, by combining the theory of crime as a socially constructed concept with the gendered lens proffered by feminism, we expose a further dimension to that of class and race. In the 1970s, Raewyn Connell interrogated gender as a form of power relationship, with violence against women as a means of their oppression.[15] In the past men have held, and continue to hold, the lion's share of economic and political power, which they have used to maintain the patriarchal structure of British society. There are at least two processes by which this has been achieved. Firstly, in a hierarchical society, a class system emerges and within each, we observe the presence of a

[11] W J Chambliss and R B Seidman, *Law Order and Power* (1971) 7.
[12] *Ibid.*
[13] Chambliss and Seidman (n 11) 65.
[14] L Gelsthorpe, "Feminism and criminology" in M Maguire, R Morgan and R Reiner (eds), *The Oxford Handbook of Criminology* (2002) 112 at 116.
[15] J Steans, *Gender and International Relations* (2013) 25.

hegemonic male frame. Through a biased criminal justice system, elite hegemonic masculinities successfully oppress opposing hegemonic masculinities. Secondly, crimes against women *as women* are minimised, facilitating the pernicious oppression of women within and by society, upon which a patriarchal structure depends. In the United Kingdom, one way the elite have successfully done so is by constructing the law in such a way that crimes which pervade women's empowerment, are stripped of their gender specificity and become hidden and normalised in their enormity. Violent crimes against women are pivotal to maintaining male dominance and are indiscriminate of class or race. For both, the process remains the same. We observe political, patriarchal ideology, justified by legislation and legitimised by criminology.

E. GENDER, JUSTICE AND MORTALITY

Before analysing the gendered nature of crime and criminology, I will make a moral argument for the gendered nature of justice. Chambliss and Seidman argue that "regardless of how homogenous a society may at first glance appear, behind the cloud of consensus and unanimity there always lurks the fact of widespread disagreement on what constitutes the 'right and proper' thing to do".[16] We might then be tempted to turn to philosophy for moral guidance. As it stands, the law advocates a Kantian ethical code of moral absolutes. However, this proves problematic, as in order to have moral absolutes, we must contend with the existence of the metaphysical manifestation of what is "good". The alternative to this is taking a utilitarian approach, where rather than a deontological logic of "means rather than ends", we instead make an assessment of "ends rather than means".[17] Specifically, utilitarianism is based on a maximisation principle of human happiness. When examining crime, we might invert this to the minimisation of harm. This idea has developed recently with the birth of zemiology, a critique of criminology which focuses on collective harm as a result of criminal acts, as opposed to the acts themselves. There has also been a study of victimology, which analyses the effects of crime for victims. I argue along this line of reasoning, that we might consider the gendered nature not only of crime, but also of justice. Popular opinion contends that justice is served through suitable punishment: when a perpetrator is quite literally, "brought to justice". However, very seldom do criminologists consider, analyse or make provisions for the victim, with the time, energy or resources equivalent to that of perpetrators. I exemplify this using paedophilia. While I have addressed the problems associated with the concept of a social consciousness, if we take media coverage as an indicator of public opinion, we could argue that paedophiles are some of the most vilified and despised individuals in society, with many relegating them to sub-human status. Yet, women in prison garner very little empathy or interest from the general population, despite the fact that, as Baroness Corston points out in the 2007 Home Office Report on women's prisons, many of these women were victims of childhood sexual abuse.[18] Therefore, I argue that punishment falls into a masculine frame, given the association with aggression, discipline, and vengeance. I argue from a feminist perspective that we should, therefore, promote a complementary "feminine" form of justice. Nurturing, mothering and caring are often classified as "feminine". Thus, were we to incorporate these equally into our conception of justice, the treatment of the victim is just as - if not more - important as the treatment of the perpetrator. With zemiology in mind (which comes from the Greek *zemia*), one might consider the Aristotelian concept of *eudaimonia*, which translates to "human flourishing".[19] On the one hand, we have harm, and the subsequent "masculine" need for perpetrator punishment, but simultaneously we should encourage and facilitate the *eudaimonia* of the victim as its "feminine" counterpart. Together, I believe this would constitute a more complete, gender-balanced form of justice. From justice, I move

[16] Chambliss and Seidman (n 11) 65.
[17] R Johnson and A Cureton, "Kant's Moral Philosophy", in *The Stanford Encyclopedia of Philosophy* (2017).
[18] J Corston, "The Corston Report: A review of women with particular vulnerabilities in the criminal justice system" (2007) at 70, available at http://webarchive.nationalarchives.gov.uk/20130206102659/http://www.justice.gov.uk/publications/docs/corston-report-march-2007.pdf.
[19] R Crisp, "Well-Being", in *The Stanford Encyclopedia of Philosophy* (2017).

to the gendered nature of crime, specifically for men.

F. HEGEMONIC MASCULINITY AND CLASS

Michalowski argues that a capitalist construction of society allows the elite to exploit the class system and use their relative power to influence the law, thereby exposing the concept of criminality as bias towards lower classes.[20] Given that ethnic minorities and those with disabilities are highly represented in lower socio-economic classes, there are further racial and ability related cleavages. Furthermore, Sheila Rowbothan argues that this capitalist relationship of property and domination is divided along gendered lines.[21] In *The Limits of Masculinity*, Andrew Tolson explores what is meant by "doing masculinity", arguing that there is no "universal masculinity, but rather a varying masculine experience of each succeeding social epoch".[22] Therefore, the elite hegemonic male, constructs the law with a bias against other forms of masculinity. Tolson elaborates on the relationship between elite hegemonic masculinity and capitalism: "it is important to a man that he can talk about his ancestors, his property [...] he can invoke the ancient law of 'patriarchy'[...] It provides a man with a reference point for self-justification [...] in the masculine gender identity".[23] However, this form of masculinity is elitist. Property is not obtainable by all, and detailed family history and ancestry is something available only to those in the middle-to-upper classes. Therefore, if this is the basis for masculinity, it is a masculinity only some might achieve. As a result, for fear of emasculation, men who cannot base their masculinity on property or ancestry, create an alternate frame using accessible means. Tolson writes that "through working, a boy [...] becomes a man, becoming part of a vast economic organisation [...] There is a magnetism in the paternal birth-right that transcends the daily monotony of wage-labour".[24] It is to the benefit of an elite male that those of lower social standing remain unchallenging to his position, and within a capitalist society, do so while remaining willing to sell their labour. Therefore, work ethic is a convenient element of working class hegemonic masculinity which exists under the threat of a biased criminal justice system, given the investor class have more power and capital to influence the law and, thus, criminology. However, I argue that it is not simply a class or racial struggle, but also a masculine one. It is the construction of an elite hegemonic masculinity, which criminalises and demonises opposing masculinities. "Criminological theory has largely worked on the assumption that crime is an overwhelmingly youthful, masculine, working class activity".[25] This is a compelling explanation for why, despite evidence indicating that being male and working class are indicators for criminal involvement, the law might be skewed to produce such a result. The first process through which this occurs, is through the criminalisation of actions most likely to be undertaken by men in lower social classes.

G. CRIME AND THE POOR

Inequalities in wealth and power must be maintained for an elite class to exist and so lower classes are prevented from obtaining either through the process of law. "The crux of the problem here is that the social class that control and most benefit from corporate capital accumulation are also the same social class that exert powerful agenda-setting pressures when it comes to law making and law enforcement priorities".[26] The most transparent way that individuals are demonised as deviant in the eyes of the law is by criminalising those acts for which they are most likely to be guilty. Given that individuals from ethnic minority backgrounds are more likely to occupy these social classes for historic reasons and reasons related to institutional racism, there is a further racial dimension to this argument. To

[20] Michalowski (n 1).
[21] S Rowbothan, *Hidden from History* (1973) ix.
[22] A Tolson, *The Limits of Masculinity* (1977) 13.
[23] *Ibid.*
[24] *Ibid.*
[25] I Taylor, P Walton and J Young, *The New Criminology: for a social theory of deviance* (1981) 15.
[26] R Michalowski (n 1) at 189.

illustrate this, I will use the example of cocaine in the United States. In America, black people are more likely to hail from lower economic backgrounds. Crack cocaine, which is relatively cheap compared to its powdered counterpart, incurs a sentence for possession of up to 25 years. White people, and those from higher socio-economic backgrounds, are more likely to use powdered cocaine, and its possession carries a monetary fine. Therefore, if a black man is caught and sentenced for possessing crack cocaine, the subsequent, disproportionate sentence punishes him not for his proximity to narcotics, but for being poor, black and male. Similarly, consider the financial crash of 2008. When sub-prime mortgages handed out over the previous ten years could not be repaid, they, and every bond made against them imploded, the government bailed out the banks at a cost of £19,721 per capita, and Britain has been in a state of austerity ever since.[27] There is no limit to the consequences which can be attributed to this, from a huge increase of children living in poverty in the United Kingdom, to higher instances of domestic violence,[28] which some believe to be behaviour associated with emasculated males as a result of job loss, as well as increased suicide rates and substance abuse.[29] In turn, these are all causal or actual instances of criminal offending, which require the involvement of the criminal justice system. However, given the banking industry was deregulated under Thatcher and even further under Blair, it was advertised as bad business, as opposed to illegal behaviour. Therefore, little to no subsequent criminological study has focused on the financial crash, despite its substantial repercussions throughout the world. Some banks were guilty, but this is a fact omitted from the common narrative. Michalowski's argument holds; deviant organisational culture, such as corporate foul play, garners far less legal or political attention than individual criminal acts.[30] It defies logic that such a significant period of time and energy is spent studying the deviant behaviour of one section of society, when the same behaviour found in elite circles has far greater consequences, yet fails to warrant criminological enquiry. Michalowski refers to this as the "criminological inversion",[31] where crimes affecting the least number of people warrant the maximum inquiry, while those committed on an institutional, elite level, warrant little to no inquiry. Were this a moral argument of either deontological absolutes or utilitarian maximisation, the elite class would be considered guilty of actions significantly more immoral than those in lower social classes. However, this is inverted by the criminal justice system, and this is legitimised by criminology.

H. WHAT WARRANTS CRIMINOLOGICAL ENQUIRY?

While I agree with Michalowski's claim that the law is created with class bias, I argue that it is as much an example of hegemonic male oppression, as class oppression. Lower social classes have unique hegemonic masculinities, which are exploited by the construction of crime. Walklate argues that young males who hail from these backgrounds have higher instances of delinquency than young females for two reasons: "[...] they are less strictly controlled by the socialisation process and […] through processes are taught to be tough, aggressive, active, risk-seekers".[32] Cohen assumed delinquency was "primarily a working-class phenomenon [...] judged by middle class values".[33] For

[27] J Stiglitz, "Austerity has strangled Britain (2017) *The Guardian*, available at https://www.theguardian.com/commentisfree/2017/jun/07/austerity-britain-labour-neoliberalism-reagan-thatcher.

[28] H A Ritchie, "Gender and enterprise in fragile refugee settings: female empowerment amidst male emasculation" (2018) 42(S1) Disasters S40.

[29] Dr S Milio et al, "Impact of the Economic Crisis on social, economic and territorial cohesion of the EU", for the European Parliament, (2014), available at http://www.europarl.europa.eu/RegData/etudes/etudes/join/2014/529066/IPOL-REGI_ET(2014)529066_EN.pdf; H Kriesi, "The Political Consequences of the Financial and Economic Crisis in Europe" (2012) 18(4) Swiss Political Science Review 518.

[30] Michalowski (n 1) at 184.

[31] *Ibid.*

[32] S Walklate, *Gender, Crime and Criminal Justice* (2004) 66.

[33] S Cohen and L Taylor, "Making out and fighting back" in S Cohen and L Taylor (eds), *Psychological Survival: the Experience of Long-term Imprisonment* (1972) 129 at 135.

both Sutherland and Cohen, sex-role theory was pivotal.[34] However, I argue that it is not sex but gender that plays a defining role in how men are taught to "do gender". As Messerschmidt argues "[...] men construct masculinities in accord with their position in social structures and therefore their access to power and resources".[35] This is then perpetuated by criminologists, for whom the two following two conditions must be satisfied to warrant inquiry:

a. an action or omission that constitutes an offence that may be prosecuted by the state and is punishable by law; and

b. an action or activity that *although not illegal,* is considered to be evil, shameful, or wrong.[36]

As Raffaele Garofalo argues, this proves to be deeply problematic: "[t]he attempt to show us what the law views as crime ends in our being told that crime, in the eyes of the law, is the doing of that which the law itself forbade".[37] Therefore, an elite class can construct a hierarchical system to maintain control of resources, defining the law with bias in order to maintain power. Michalowski writes:

> [...] the social imaginary created by legalist criminology is that, with the exception of those types of corporate behaviours explicitly criminalized by law, the social injuries and harms resulting from the corporate pursuit of profit, capital accumulation and *power* are not subjects for criminological inquiry.[38]

Patriarchal dominance within masculinity is thus justified and legitimised. While men as criminals remain front and centre of criminological inquiry, women and the crimes committed against them are relegated to the shadows.

(1) Violence against women

Radford and Stanko argue that violence is "[...] used by men as a way of securing and maintaining the relations of male dominance and female subordination central to the patriarchal social order".[39] Mackinnon's sentiments are similar: "[...] the widespread existence of men's violence against women and children is essential to a system of gender subordination".[40] Further, in their extended critical analysis of rape, Schwendinger and Schwendinger found strong links between rape, sexual assault and sexual inequality in capitalist socioeconomic systems.[41] Power is exercised over women in a pervasive, systematic and pernicious manner while remaining terrifyingly invisible. Radford writes that "[...] we live in the midst of a reign of sexist terror comparable in magnitude, intensity, and intent to the persecution, torture, and annihilation of European women as witches".[42] Women exist with an oppressive fear of violence. It is a dynamic created by a patriarchal society and upheld by the law of the land. The first way I argue that this occurs is that the law hides crimes against women by denying their gender specificity.

---

[34] R W Connell, "Theorising gender" (1985) 19(2) Sociology 260.
[35] J Messerschmidt, "Men, Masculinities and Crime" in M Kimmell, J Hearn and R Connell (eds), *Handbook of Studies on Men and Masculinities* (2005) 238.
[36] Michalowski (n 1).
[37] R Garofalo, *Criminology* (1914) 59.
[38] Michalowski (n 1).
[39] J Radford and E Stanko, "Violence Against Women and Children: the contradictions of crime control under patriarchy" in K Stenson and D Cowell (eds), *The Politics of Crime Control* (1991) 186 at 186.
[40] C MacKinnon, *Feminism Unmodified: Discourse on Life and Law* (1987) 67.
[41] J R Schwendinger and H Schwendinger, *Rape and Inequality* (1983).
[42] J Radford, *Femicide: The politics of woman killing* (1992) 20.

(2) Hidden violence

Domestic violence, but more critically the fear of violence, is gendered. While domestic violence can be experienced by both sexes, in 2016 women were twice as likely to experience intimate partner violence in England and Wales.[43] The most prevalent form of abuse for men was stalking, whereas women were five times more likely to experience sexual assault. However, the rhetoric used to document violence against women disguises the gendered nature of such crimes. Analysis of men's violence against women becomes a gender-neutral problem of "family violence".[44] Therefore, while the law recognises marital problems and violent problems, the problem of men exercising power over women, violently or otherwise, is hidden. One way feminists have tried to uncover the gendered nature of crime is by redefining the murder of women as "femicide". As opposed to homicide, they define "femicide" as: "the killing of a woman or girl, in particular by a man, and on account of her gender".[45] In 2016, of the 113 women murdered in England and Wales, 69% were killed by their partner or previous partner.[46] A further 7% were killed by their son or another member of the family.[47] This is four times the number of men who were killed in a domestic violence context.[48] However, when domestic violence results in death, it is documented as a homicide issue, not an issue of domestic violence against women. To some, this seems a tenuous claim. However, if we consider racially motivated or homophobic homicide, most people accept the distinction and specificity between these, and other homicidal acts. Rhetoric is powerful, and denying the gendered nature of violence against women is pervasive. Women's fear of violence was found by the British Crime Survey to be out of all proportion with their statistical chances of victimisation.[49] In turn, this is trivialised by the establishment, with the authors of the survey labelling this fear "irrational".[50] Even this choice of word is gendered and has numerous negative connotation of the "irrational woman". Rhetoric, media and culture within a patriarchal society advertise the prevalence of violence against women, while simultaneously concealing it as gender-neutral. A huge part of this process is the manner in which violence against women is portrayed in public narrative, such as media coverage.

(3) Violence against women in the media

Propaganda which shows heinous violence against women is widespread. As Stuart Hall writes: "[c]rime news is news".[51] However, combined with a concealed narrative of gender-based crime, women's fear of violence appears irrational. The journalistic narrative of crimes against women is gruesome, titillating and dehumanising. Furthermore, headlines and stories describing the heinous acts in detail often refer to parts of the woman's body in a manner which is objectifying. In July 2017, a nineteen-year-old woman was murdered. The Metro ran the headline "Body of girl, 19, found chopped up in freezer" and in the article, the crime is graphically detailed.[52] Further, murdered women are frequently cast in a particular role by the media, from Madonna, to whore, to monster, with sympathy elicited from news outlets often dependant on the role cast. A number of newspapers thought it necessary to disclose that the aforementioned murdered woman was a virgin, perhaps then beyond all doubt undeserving of the crime. A further example is the murder of Reeva Steenkamp at the hands of her boyfriend Oscar Pistorius. The story was front-page news and newspapers overwhelmingly ran the story with a modelling picture of Reeva, in her underwear, looking suggestively out at the reader. Headlines shrieked "Blade Slays Blonde", "3 Shots. 3 Screams.

[43] Office for National Statistics, "Intimate personal violence and partner abuse", A Crime Survey for England and Wales (2016).
[44] Radford and Stanko (n 39) at 187.
[45] Radford (n 42) 3.
[46] Women's Aid, "Until women and children are safe", Femicide Census (2017).
[47] *Ibid.*
[48] NHS, *Domestic Violence London* (2017).
[49] See the discussion in J Radford and E Stanko (n 39) at 192 *ff.*
[50] *Ibid.*
[51] S Hall et al, *Policing the Crisis* (1978) 66.
[52] The Metro, "Body of girl, 19, found chopped up in freezer" (2017), available at https://metro.co.uk/2017/07/21/body-of-girl-19-found-chopped-up-in-freezer-at-home-in-kingston-6795731/.

Silence. 3 More Shots"[53] and "Pistorius DID beat model girlfriend with a cricket bat".[54] I argue that Reeva was cast in the role of the whore, and sensationalising her murder denied the gravity of her murder. This technique is reminiscent of slasher films or Grand Theft Auto-style prostitute assassination.[55] At best, it is using a woman's death to sell papers, but at worst, there is an underlying, undeniable, suggestion that her sexuality - how she was "doing gender" - in some way made her deserving of the crime. The subsequent trial of Pistorius displayed him with far more humanity than that awarded to Steenkamp. Photos of Pistorius weeping in court were widespread, as were tales detailing the toll the trial was taking on his well-being. He was sentenced to six years in prison. The story was unavoidable, and the message to the women of the world was clear. Six years for six shots in cold blood.

Just as women are cast in a particular role and frequently objectified, Pistorius is not unique in his coverage. The humanity and normality of male perpetrators is frequently portrayed in news stories, often through quotes from friends and neighbours. For example, in 2016 Lance Hart waited in a car park for his wife and daughter for three hours with a shotgun before shooting them both and turning the gun on himself. A neighbour was quoted as saying that Hart was "the nicest guy you could ever meet […] he would do anything for anyone".[56] Crimes against women are minimised both institutionally and by the media, whereas male perpetrators are shown in a more favourable light. For example, in the United States, the FBI terms sex killings "recreational murder".[57] In Canada, Marc Lépine, smarting from his rejection from École Polytechnique in Montreal, and angered by women holding positions traditionally occupied by men, burst into a college classroom, ordered the men out and opened fire. Fourteen women were killed and ten were injured. Lépine's suicide note was found after the massacre. It stated that: "[…] feminists have always enraged me. They want to keep the advantages of women (e.g. cheaper insurance, extended maternity leave preceded by a preventative leave, etc.) while seizing for themselves those of men".[58] However, following the act of mass femicide, Quebec's Prime Minister Robert Bourassa denied that the crimes had been motivated by gender and "[...] rejected petitions to close the legislature [...] as a day of official mourning was only appropriate, he insisted 'when someone important to the State has died'".[59] Therefore, we witness crimes against women being treated as though *less criminal* on a political, institutional, and cultural level, which oppressively permeates each and every woman's subconscious. The fear is dismissed as irrational, but it is all too real, and it is supported, justified and legitimised by the law.

(4) Lip-service law

The law legitimises the political desire to oppress women, supporting the message portrayed to women that crimes against them are of little consequence. I analyse three means by which this occurs; (1) through the law, (2) by lawyers and (3) through potential jury bias. Since women's suffrage, we can track the slow progress of incorporating crimes against women into common law. However, rape and rape tolerant attitudes are still prevalent, and rapists continue to be given shockingly lenient sentences. Chambliss and Seidman offer a compelling argument for "lip-service law":

> It seems to be the case that two things characterize laws passed which superficially seem inimical to the best interests of the powerful. First, some laws are passed primarily as a token

[53] See discussion of both headlines by the Media Diversity Institute, available at http://www.media-diversity.org/en/index.php?option=com_content&view=article&id=2439:coverage-of-reeva-steenkamps-murder-shows-media-sexism&catid=35:media-news-a-content&Itemid=34.
[54] The Mail on Sunday, "Pistorius DID beat model girlfriend with a cricket bat" (2013), available at http://www.dailymail.co.uk/news/article-2283507/Oscar-Pistorius-DID-beat-model-girlfriend-Reeva-Steenkamp-cricket-bat-shot-her.html?cmp.consent=true.
[55] Radford (n 42).
[56] The Sun, "Triple Shooting Tragedy" (2016), available at https://www.thesun.co.uk/news/1467233/pictured-daughter-after-gunshot-is-heard-at-lincolnshire-swimming-pool/.
[57] Radford (n 42).
[58] M Kimmel, *Angry White Men: American Masculinity at the End of an Era* (2013) 84.
[59] Radford (n 42) 47.

> gesture to assuage a group that has the capacity to disrupt the ongoing processes of society. Secondly, laws which are passed under these conditions are rarely ever enforced with the same vigour as characterises those laws which are in the best interests of persons in positions of power.[60]

If we consider the most powerful in society to be hegemonic males who rely on the oppression of women for the maintenance of patriarchy, it stands to reason that laws criminalising oppressive violence against women will be introduced as token gestures and less vigorously upheld. For example, in the United Kingdom prior to 1991, rape was not considered a criminal act within the institution of marriage, with many arguing that sex is "the conjugal right of the husband".[61] Meanwhile, feminist activists worked tirelessly for its criminalisation, while being met with public ridicule and anger. They fought to end the belief that "when a woman says 'I do' she gives up her right to say 'I don't'".[62] In response to the movement, divorce lawyer Sidney Siller penned an article in 1983 claiming "it would only be a slight exaggeration to say a man has a better chance of surviving a nuclear holocaust than winning a legal contest with a woman" in an article eloquently entitled: "Wife Rape- Who Really Gets Screwed".[63] Legalised rape within marriage is an example of the law legitimising the patriarchal belief that a woman as a wife constitutes a possession. Similarly, this is tied to the concept of property as an element of masculinity. It is an element found in most hegemonic male frames, as a wife may be the property of all men, elite or poor. Thus, for each frame, even when the possession of material goods is unobtainable, there is the possession of a wife. Therefore, the patriarchal ideology of wife possession is enforced and justified by law. In fact, rape and abuse within marriage often has more to do with maintaining patriarchy within the family structure than about sex or violence. As Mackinnon writes:

> Physical battering, [in the eyes of the criminal justice system] is either a reflection of bad marital relations, personal disputes, or intoxicating substances, not the manifestation of unequal power and a need for control. Sexual abuse, following the same line, arises because of some men's uncontrollable lust or miscommunication with women and children, not as an exercise of patriarchal power.[64]

Thus, the gendered nature of crimes against women are denied.

(5) Victim on trial

Another prominent tactic used to delegitimise crimes against women is that of "victim participation", achieved by assuming a level of women's active participation in the crimes committed against them. Regarding rape, interactionist theorists such as Menachem Amir make a case for rapist and victim being locked in a relationship as "mutually interacting partners".[65] A woman might use "indecency in language and gestures […] did not resist strongly enough […] [or] agreed but retracted".[66] This is a theme so prevalent in rape rhetoric that campaigns aimed at rape reduction appeal to women to keep safe, not to men to better understand consent or simply to not rape. As a result, women who are raped are questioned on their attire, alcohol consumption, how hard they struggled or whether they simply had morning-after regret. When alcohol was consumed, or clothing considered risqué, there is an insinuation of having been "asking for it". Furthermore, victims of "femicide" are dragged posthumously through the mud. Their character and humanity are desecrated with questions of what they might have done to deserve the attack.[67] Arguments are formulated in defence of the perpetrator

[60] W J Chambliss and R B Seidman (n 11) 65.
[61] D E H Russell, *Rape in Marriage* (1980) xi.
[62] *Ibid.*
[63] S Siller, "Wife Rape - Who Really Gets Screwed" (1983) 14(9) Penthouse 104.
[64] C MacKinnon, *Feminism Unmodified: Discourses on Life and Law* (1987) 67.
[65] J R Schwendinger and H Schwendinger, *Rape and Inequality* (1983) 65.
[66] *Ibid.*
[67] J Radford (n 42).

regarding the sexual preferences of the victim or how well she performed her role as a wife or mother. Performing either of these roles inadequately or being overly sexualised are commonly used to reduce the perceived severity of the murder committed. The point I find most paradoxical regarding the victim participation narrative is how contrary it is to the point I made earlier about female criminals. Very seldom is the "victim participation" narrative used as evidence for women who are sexually abused and robbed of innocence by paedophiles. Instead, when they are later involved in criminal activity, they are demonized by the state and the general public, with a limited narrative regarding the role played by their attacker. Women may play a role in crimes committed against them, but men may not play a role in a woman's justification for her crimes. Along with the media, lawyers can then be found to perpetuate this narrative and, therefore, the potential for jury bias is clear. The media portrayal of crimes against women permeates each adult individual in society, a biased formulation of laws benefits elite males and maintains the hegemonic order of women as possessions, resistance is played out publicly and ridiculed, and thus patriarchy becomes further ingrained in societal norms.

I. CONCLUSION

In societies such as the United Kingdom, we witness patriarchal political ideology justified by the law and legitimised by criminology. Therefore, I have argued that the law should always be contextualised by criminologists, taking into account the political and economic landscape. Viewing the law within its gendered context allows us to expose the manner in which it is biased against men based on class, and against women more generally. A gendered conception of justice together with lip-service law allows violence to be exerted as a form of social control. The media titillates and sensationalises this violence, causing women to live in fear. The victim-participation narrative further demonises female victims, assigning a level of responsibility to them, for the crimes committed against them. Ideology is justified by law. Furthermore, criminology legitimises this process if contextually uncritical, as I have argued that criminological enquiry cannot be value neutral. Indeed, lacking objectivity is not necessarily a threat to the field. However, criminologists must be aware of the lack of objectivity. What a feminist perspective offers to this awareness is a gendered dimension. Reframing the questions we ask is vital, not for an objective practice, but for a practice aware of its subjectivity. Why did he defraud the benefit system? *Why were the banks allowed to defraud the country?* Why didn't she say no? *Why didn't he ask her if she was saying yes?* Why do women stay with abusive men? *Why do abusive men not let women leave?* These are the questions criminologists should be asking in an attempt to escape patriarchal bias and more accurately, and critically, study criminal behaviour and deviance.

CRITICAL CRIMINOLOGY: ITS CURRENT MARGINAL STATUS QUO

Tong Lu[*]

A. INTRODUCTION

Crimes, perceived as a social phenomenon, have frequently been the subject of heated discussion. Disparate from criminal law, which focuses more on punishments and rehabilitation, criminology tends to explain the rationale behind criminal behaviour and its future prevention. Numerous experts on crimes and criminal justice systems have adopted distinct methodologies to undertake this task. For instance, some criminologists have focused on the intrinsic nature of human beings, arguing that some people are born with the urge to commit crimes.[1] Others discuss the interrelations between criminals and their outer environment, theorising that the reasons why people commit crimes can be traced back to their living environments. Another insightful approach to criminology concentrates on changes in social, political and economic circumstances, suggesting a causal link between those changes and the variation of the crime rate.

Each theory within criminology undoubtedly provides a level of clarity regarding the causes of crime. Nevertheless, critical criminology, focusing neither on the inherent nature of human beings nor criminals' external environments, offers a particularly insightful perspective. It holds a critical view of all previously existing theories within criminology, accusing them of ignoring the whole social structure and its relationship with crime. This critical criminology, also referred to as radical criminology, addresses the deeply rooted inequality in society caused by unbalanced economic distribution and exploitation. Its revolutionary value is further highlighted in its rejection of readily defined concepts, such as crime and deviance, and its emphasis is on the actual criminalisation of behaviours.[2] Insightful as it is, due to its inherent flaws and the emergence of further new approaches to criminology, critical criminology is arguably no longer the main theory in accommodating current criminal conditions and systems. In this article, the emergence and developments of critical criminology will be outlined first. The value of the theory will next be discussed, while illustrating its main contentions and consensus. The third part of the article will be dedicated to explaining the

[*] LLM student in International Law at the University of Edinburgh. This article is an adapted version of an essay presented for the LLM course "Theoretical Criminology".

[1] C Lombroso, *Crime: its Causes and Remedies* (1994).

[2] W S DeKeseredy and M Dragiewicz, *The Routeledge Handbook of Critical Criminology* (2012).

rationales behind critical criminology's current marginal status. This article will conclude by suggesting that, despite being marginal in general, the core of critical criminology, which is its critical perspective and attention on the inequality of society, is indeed at the heart of many new approaches to criminology.

B. BRIEF HISTORY OF CRITICAL CRIMINOLOGY

Critical criminology initially emerged in the late 1960s to early 1970s. The main reason for its development was dissatisfaction regarding the existing mainstream approaches to criminology, including classical criminology and positivist criminology.[3] Classical criminology focuses on the inherent nature of individuals, suggesting that criminals are in fact acting in adherence to their free will, irrelevant of any forces.[4] It is asserted that the cause of crime lies in the individuals' desire to maximize pleasure and minimize pain.[5] Consequently, committing crimes is a well-calculated exercise after weighing up the costs and gains. Alternatively, positivists suggest various specific factors that might contribute to criminal behaviours. For example, there might be biological reasons, psychological standings or social backgrounds that can lead to the perpetration of crimes.[6] It follows that individuals are not necessarily choosing to commit crimes. Instead, they are driven by different forces to conduct criminal behaviours. Despite the divergent explanations, both mainstream theories focus on the correlations between individuals and crime-commission. The causes of crimes, according to those two approaches to criminology, can be readily traced back to individuals.

This absolute concentration on individuals was criticised by United States (US) sociologist Howard Becker,[7] the accepted founder of labelling theory, which greatly influenced the development of critical criminology. In an attempt to explain the existence of crime, Becker focussed on the social reactions to behaviours, which can create further deviance. He stated that once a person violates one of the rules designed by some groups of people in a society, that person will be labelled as a deviant. Moreover, because of the social reactions to people who have been tagged as troubled, or even criminal, those people might have a higher chance of committing an actual crime, or a more serious one.[8] The reason behind this reaction is two-fold. Firstly, society may stigmatize that particular person for their wrongdoing and exclude him or her as a result. Being isolated can ultimately lead to the commission of a crime. Secondly, individuals who have been labelled may unconsciously be more prone to living up to that label, suggesting that they may vary their future paths by viewing themselves as deviants.[9] The main contribution of labelling theory, which has also had a vital impact on later critical criminology, is the shift of attention from crime and criminal to a social mechanism beyond the phenomena. Despite that contribution, labelling theory has nonetheless been denounced as merely focusing on the micro-level of offence, while failing to relate the phenomena of deviance to broader social, historical, political and economic contexts.[10] Critical criminology, as will be discussed, advances more macro-level factors that may also explain crime and deviance, for instance, the power of a ruling class and inequality within a society.[11] The inescapable impact of Marxism arguably swayed such contentions.[12]

---

[3] D Downes and P Rock, *Understanding Deviance* (2011).
[4] S Cohen, *Against Criminology* (1988).
[5] J Bentham, "A Comment on the Commentaries and A Fragment on Government", in J H Burns and H L A Hart (eds), *The Collected Works of Jeremy Bentham* (1977) 393.
[6] R Burke, *An Introduction to Criminological Theory*, 3rd edn (2009); W C Reckless, *Vice in Chicago* (1933).
[7] H Becker, *Outsiders: Studies in the Sociology of Deviance* (1963).
[8] E Lemert, *Human Deviance, Social Problems and Social Control*, 2nd edn (1972).
[9] P Rock, "Sociological Theories of Crime", in M Maguire, R Morgan and R Reiner (eds), *Oxford Handbook of Criminology*, 5th edn (2012).
[10] A Liazos, "The poverty of the sociology of deviance: nuts, sluts and perverts" (1972) 20 Social Problems 104.
[11] I Taylor, P Walton and J Young, *The New Criminology for a Social Theory of Deviance* (1973).
[12] R Heilbroner, *Marxism, For and Against* (1980) 15.

Some scholars have credited the United Kingdom (UK) and the US as the birthplaces of critical criminology,[13] while others have proposed European countries as its origin.[14] Regardless of the location, the social context under which the theory has developed is predominantly capitalist, where the owner of the property has paramount status. Although Karl Marx said little about crime *per se*, many critical criminologists base their theories on Marxist perceptions of capitalist society, in that the notions underlined correctly fit the existing society.[15] From a Marxist standpoint, the law is "a part of a superstructure adapting itself to the necessities of an infrastructure of productive forces and productive relations".[16] Foundational critical criminologists have used this class-division concept of Marx to explain crime. It is suggested that, in essence, the law is an instrument for the ruling class to unfairly distribute resources and labour-powers and to consolidate its existing power. Ultimately, crimes are products of defective capitalist social construction, which are mainly caused by conflicts between the two classes.[17] According to critical criminologists, the individual, instead of being a deprived underdog, is a part of an underclass fighting back and the criminal justice system is "a politically economy of criminalization".[18]

C. REVOLUTIONARY AND INSIGHTFUL

Through rejecting existing approaches to criminology, critical criminology is deemed revolutionary in several ways. Due to a lack of precise definition,[19] the contentions and crucial tenets of the theory will be analysed, together with its insightfulness.

(1) Anti-essentialism and social construction

The emergence of critical criminology was revolutionary, firstly by fundamentally rejecting the previously existing mainstream approaches to criminology. It explicitly criticised the concepts that previous criminologists had taken for granted and suggested a different approach to understanding and defining terms such as crime and deviance. In this sense, critical criminology is arguably anti-essentialist. Anti-essentialism is a philosophical concept embracing the belief that there are no specific traits or characteristics that must be possessed by certain entities for them to be "the entities".[20] It means that the entities do not have any intrinsic qualities. Critical criminologists, adopting anti-essentialist ideas, posit that before categorising the cause of crime, it is a prerequisite to determine what constitutes a crime, namely, the criminalisation of specific behaviours.[21] According to critical criminologists, those crime-related concepts are essentially social constructions which do not exist either ontologically or objectively.[22] For example, homosexuality, which was illegal in the UK until 1967, is legal now.[23] A similar case would be the prohibition of alcohol in January 1920 in the USA. The legitimisation of selling recreational cannabis in the state of California can further demonstrate the non-essentialism of the concept of crime. As such, some acts or behaviours have unique features that were considered criminal at one time yet can be redefined as legal in another.

---

[13] DeKeseredy & Dragiewicz, *Routeledge Handbook* (n 2).
[14] R van Swaaningen, "Reclaiming critical criminology: Social justice and the European tradition" (1999) 3(1) Theoretical Criminology 5.
[15] S Russell, "The continuing relevance of Marxism to critical criminology" (2002) 11(2) Critical Criminology 113.
[16] E P Thompson, "The rule of law in Marxism and law", in P Beirne and R Quinney (eds), *Marxism and Law* (1982) 130.
[17] J Young, *The Exclusive Society* (1999).
[18] van Swaaningen (n 14) at 11.
[19] DeKeseredy & Dragiewicz, *Routeledge Handbook* (n 2).
[20] R Patel and D Tyrer, Race, *Crime and Resistance* (2011).
[21] W J Chambliss and R Seidman, Law, Order and Power, 2nd edn (1982).
[22] Taylor et al, *The New Criminology* (n 11); L H C Hulsman, "Critical criminology and the concept of crime" (1986) 10 Contemporary Crises 21.
[23] Marriage (Same Sex Couples) Act 2013, for England and Wales; Marriage and Civil Partnership (Scotland) Act 2014, for Scotland.

Another point supporting the idea of the anti-essentialism argument is the fact that some crimes are neither against moral values nor harmful in any way.[24] Thus, there appears to be no other explanation for the criminalisation of some behaviours than to suggest that the concept of crime is socially construed. For instance, while one can suggest that spitting in the street is not in any way immoral, it was nonetheless defined as a crime in the London borough of Enfield from December 2013.[25] Such social construction can also be demonstrated by the example of the non-criminalisation of harmful behaviour. For example, killings in war are not construed as a crime, though they have the same catastrophic effects as murdering someone, whilst the latter is punished.[26] Lastly, damages to the environment by large corporations are not fully regulated by criminal justice systems, or at least, not as rigorously as they could be.[27] As the aforementioned examples suggest, crime is not precisely linked with morality or harmfulness, but can be a product of social construction and might also be subjected to revision, periodically, within a dynamic society.

(2) Capitalist and ruling class domination

Another valuable insight provided by critical criminology is the change in focus from the identity of the criminal to the identity of those criminalising behaviours. After rendering the concept of crime as a social construction, critical criminologists went on to assert that it is the authoritative class in society that dominates such a construction.[28] This idea is readily inherited from the Marxist perspective, which separates community into two groups: the ruling class and the working class. It is proposed that the former, which has supreme power and resources, can radically and fundamentally affect and even control the law.[29] This proposition can be illustrated by an example from US history: the Nixon administration. In 1970, the Comprehensive Drug Abuse, Prevention and Control Act was introduced to control the distribution and the use of dangerous drugs.[30] The drugs that fell into hazardous categories prohibited by the Act included imported narcotics and those that could be self-produced. It did not include pharmaceutically manufactured drugs, which were frequently sold illegally and caused harm. Although experts had testified against the notorious nature of some drugs produced by pharmacies, claiming that they could be harmful to the public, the US Congress did not put them on the prohibited list. This Act axiomatically demonstrated the ability of dominant groups to shape legislation in line with their interests and to shield any interference with their interests.

Moreover, it must be acknowledged that the Act enabled law-enforcement agents to enter private dwellings without permission to search for, and investigate, the possible existence of the prohibited drugs. This empowerment ultimately led to easier arrests and convictions of drug-users who were not taking drugs manufactured by pharmaceutical companies. The powerless class is, as a result, more easily criminalised, whilst the authoritative class remains free from interference with their interests. In other words, the ideas and concepts of criminal law are indeed construed through the interests and prejudices of the people who create, or who can significantly influence them. After providing definitions of crime and deviance, critical criminologists suggest that it is the ruling class that established the criminal law. By defining the exact scope of crime based on their interests, the powerful can freely criminalise behaviours that are more prone to be conducted by the working class.

---

[24] J Young, "Critical Criminology in the Twenty-First Century: Critique, Irony and the Always Unfinished", in K Carrington and R Hogg (eds), *Critical Criminology: Issues Debates, Challenges* (2002) 251.
[25] Enfield Council, "Spitting to be Banned in Enfield" (21 November 2013), available at *https://www.enfield.gov.uk/news/article/907/spitting_to_be_banned_in_enfield*.
[26] International Committee of the Red Cross, Geneva Convention Relative to the Protection of Civilian Persons in Time of War (Fourth Geneva Convention), (12 Aug 1949) 75 UNTS 287.
[27] F Pearce, *Crimes of the Powerful* (1976).
[28] S Box, *Power, Crime and Mystification* (1983).
[29] Chambliss & Seidman, *Law, Order and Power* (n 21).
[30] Comprehensive Drug Abuse Prevention and Control Act of 1970, Pub L No. 91-513, 84 Stat 1236 (27 October 1970).

In the same vein, it is not surprising to find that most of the heavily controlled crimes are petty ones, which are not fundamentally damaging to the order or moralities of a society.[31] Instead, more serious crimes, or at least more financially damaging ones, such as corporate crimes or white-collar crimes, whereby numerous innocent people are embroiled and harmed, are less criminalised and punished.[32] Within such a construction, where class bias is involved in the process of legislating, stereotypes can additionally be formed that might ultimately stigmatise the working class, causing more damage to them. It is this focus on the criminalisation initiated by influential people or groups that is considered one of the most significant contributions of critical criminology.[33]

(3) Class struggle leading to crimes

After illustrating that concepts of crime are socially constructed by the ruling class, critical criminologists further produced an insightful notion regarding the causes of crime. According to them, society is stratified and unequal between the ruling class and the working class in that the former is exploiting the latter by using their labour-power for their benefit. The class conflict is further illuminated by the fact that there is never a fair or just distribution of resources in a capitalist society. This inequality is arguably permitted and even supported by the state, dominating with significant power and choosing to protect the dominant class instead of the exploited underclass.[34] Following this, it is suggested that crimes are the response to the inequality between the two aforementioned classes and capitalist conditions.[35] Crimes, substantially, are ways in which the proletariat fights back, in trying to oppose the controlling power of states and the bourgeoisie. Rather than focusing on the individual criminals, critical criminologists are concerned more with the macro-level of the societal structure, and its responses to crime, together with the consequences of those responses. Instead of viewing criminals as dysfunctional human beings, critical criminologists instead consider them as victims of a partisan society. By having sympathy for the offenders, who are arguably forced to commit crimes, critical criminology offers a peculiar insight into the nature of criminals.

D. RATIONALES BEHIND ITS MARGINAL STATUS QUO

Insightful as it is, critical criminology has been subject to various criticisms regarding its main contentions as well as its effectiveness, which further leads to the marginalisation of its position within criminology.

(1) Deviance and social construction

As discussed earlier, critical criminologists reject the mainstream account of deviance, instead proposing that it occurs because of the resistance of the underclass against the exploitative ruling class. That said, by insisting on the paramount power of a particular ruling class in society, the notion of deviance can seemingly be defined as a way for the ruling class to exploit the underclass and maintain bourgeois hegemony.[36] These two disparate explanations of the concept of deviance appear to be irreconcilable, as to whether crime is artificially created by the ruling class to control the working class, or if it has to emerge because the working class is fighting against the ruling class. This

[31] P Hillyard and S Tombs, "From crime to social harm" (2007) 48(1-2) Crime, Law and Social Change 9; Hulsman (n 22) at 63.
[32] W J Chambliss, "Toward a political economy of crime" (1975) 2(2) Theory and Society 149; Box, *Power, Crime and Mystification* (n 28).
[33] M J Lynch, Critical Criminology: Oxford Bibliographies Online Research Guide (2011), available at http://criminology.oxfordre.com/view/10.1093/acrefore/9780190264079.001.0001/acrefore-9780190264079-e-246?rskey=l5JVcW&result=1; A Alvesalo and S Tombs, "Working for criminalization of economic offending: contradictions for critical criminology?" (2002) 11(1) Critical Criminology 21.
[34] P Ugwudike, *An Introduction to Critical Criminology* (2015).
[35] R Quinney, *Class, States and Crime: On the Theory and Practice of Criminal Justice* (1980).
[36] J Young, *Crimes of the Powerful: Marxism, Crime and Deviance* (1976).

irreconcilability consequently results in a circular argument, which is arguably endemic in critical criminology, with no precise definition of what constitutes crime or deviance, thereby creating confusion.[37] This dilemma, nonetheless, is contentiously solved by suggesting that deviance is a quality of act, which can be construed as a collective societal ideology.[38] It was stressed that the action that is socially unacceptable could be considered as deviant, regardless of the underlying motivation for performing that action. Ultimately, the law is not neutral, and it is utilised for controlling and preventing socially harmful actions.[39]

Pinning the explanation of deviant on an inherent societal ideology, however, seems to be contradictory to the pre-eminent claim of critical criminology, especially the non-intrinsic quality of the notion of deviant. Following that explanation of deviant, critical criminologists seem to have returned to Durkheim's theory of indicating how crime is inevitably embedded in society, and created following social reactions to particular behaviours.[40] The argument that the concept of deviance is constructed through a collective social consensus, seemingly leads to the conclusion that deviance is inherent in society. This interpretation, in turn, appears to fit the definition of essentialism, to which critical criminologists strongly object. However, it can be argued that the so-called social consensus, existing within a society, is a product or a concept implanted by the ruling class. In the modern society, people are more or less dependent on the mass media to get an opinion of the society that they live in, especially concerning criminalised activities.[41] The media, greatly influenced by the ruling class in promoting its agenda, create a generally disingenuous and unnoticeably false public consciousness.[42] For instance, the prevalence of news reports on brutal murders and cruel burglaries seems to cause moral panic among citizens, thereby rendering "tough on crime" propaganda and crowded prisons more acceptable. Defining deviance as a social consensus will still reconcile with critical criminology's primary contention that law is a social construction manipulated by the economically and politically advantaged group.

It is, however, not always the reality that the crowd will unselectively accept whatever they are fed.[43] People will no longer merely accept opinions and rules without reflecting on their truthfulness. It is advanced that one of the essential reasons for a law to be a discipline lies in its display of impartiality from mass manipulation and a sense of fairness.[44] This assertion is, indeed, bolstered by an ever-increasing number of laws passed against white-collar crimes, State crimes and so forth.[45] Critical criminologists, nonetheless, might argue that it is not entirely illogical considering the more democratic and inclusive nature of modern society. The democracy in society is, however, merely the hegemonic fraction of the interests of the capitalist; the actual performance of it is limited.[46] All such legislation, from a radical perspective, moreover, is merely for appearances with no substantial enforcement behind them.[47]

That said, the under-prosecution of the crimes of the powerful can sensibly be attributed to a lack of financial resources and the difficulty in accumulating evidence for a conviction, instead of

[37] D Dwones and P Rock, *Understanding Deviance*, 5th edn (2007).
[38] Taylor et al, *The New Criminology* (n 11).
[39] R Quinney and J Wildeman, *The Problem of Crime: A Critical Introduction to Criminology*, 2nd edn (1977).
[40] E Durkheim, *The Rules of the Sociological Method* (1938).
[41] Hulsman, (n 22) 63-80.
[42] J L McMullan and M McClung, "The media, politics of truth, and the coverage of corporate violence: the Westray disaster and the public inquiry" (2006) 14(1) Critical Criminology 67.
[43] C Sumner, "Marxism and Deviancy Theory", in P Wiles (ed), *The Sociology of Crime and Delinquency in Britain* (1976).
[44] *Ibid.*
[45] J Tierney, *Key Perspectives in Criminology* (2009); S Tombs and D Whyte, *Unmasking the Crimes of the Powerful: Scrutinizing States and Corporations* (2003).
[46] Young, *Crimes of the Powerful* (n 36).
[47] S Tombs, "Corporate Crime", in C Hale, K Hayward, A Wahidin and E Wincup (eds), *Criminology* (2013) 227.

outright neglect.[48] Crimes committed by the ruling class and States are, additionally, more readily dealt with through the growing numbers of enforcement methods available at the international level. For example, the Human Rights Committee is in place to scrutinise States' performance regarding their obligations under different human rights treaties, preventing crimes of genocide, unreasonable deprivation of people's property and so forth.[49] Together with the globalisation effect, more strident supervisions will be exercised among States, meaning that powerful crimes can no longer be ignored.[50] Although saving the definition of deviance, the social construction notion proposed under critical criminology does not seem entirely convincing in the current world.

(2) Underclass fighting back

The explanation for the cause of crime advanced by critical criminologists is also subject to criticisms. Finding a causal link between class struggle and the commission of crime seems to suggest that class distinction is a dreadful creation. As defined by critical criminologists, "class structure is an expression of the struggle of classes, the antagonistic relationship between those who own and control the means of production and those who do not".[51] Social class, nonetheless, is not utterly negative, considering its capability to serve as an inspiration and authority for the whole society. Moreover, in modern society, social class is viewed more as an ideology instead of a social reality.[52] As such, the separation and division of social classes is not as straightforward as it was in the era of Marx. Consequently, the explanation offered by critical criminology does not appear to be closely related to current crime problems.[53] The direct link between crime and class conflicts was additionally criticised as merely focussing on the externality of class to law while failing to identify the class relations epitomised in the law *per se*.[54] Social structures under capitalism are far more intricate than described by critical criminologists.[55] By endorsing a monophonic explanation for the cause of crime, critical criminology fails to recognise the conflicts between property owners as well as those among the underclass.[56] Moreover, different cultural backgrounds,[57] and the inequality among races and genders within society can also be the causes of crimes in capitalist countries.[58]

Furthermore, critical criminology has been criticised as creating an inaccurate impression because it romanticises the offender. The law-breakers, from the standpoint of critical criminologists, are like "amateur Robin Hoods" in "righteous attempts to redistribute wealth".[59] By fixing the cause of crime in the inequality among classes, critical criminologists tend to justify lawbreaking by the so-called underclass to portray them as victims in an oppressive society who "are already brutalised by the conditions of capitalism".[60] It must be acknowledged that the crimes supposedly committed by the

[48] P Almond, *Corporate Manslaughter and Regulatory Reform* (2013)
[49] United Nations General Assembly, International Covenant on Civil and Political Rights, 16 December 1966, 999 UNTS 171; United Nations General Assembly, International Covenant on Economic, Social and Cultural Rights, 16 December 1966, 993 UNTS 3.
[50] C W Mullins, D Kauzlarich and D Rothe, "The international criminology court and the conflict of state crime: prospects and problems" (2004) 12(3) Critical Criminology 285.
[51] Quinney, *Class, States and Crime* (n 35) 67-68.
[52] C B Klockars, "The Contemporary Crises of Marxist Criminology", in J A Inciardi (ed), *Radical Criminology: The Coming Crises* (1980) 92 at 97.
[53] T J Bernard, "Distinction between conflict and radical criminology" (1981) 72 Journal of Criminal Law & Criminology 362.
[54] S Picciotto, "The Theory of the State, Class Struggle and the Rule of Law", in P Beirne and R Quinney (eds), *Marxism and Law* (1982) 169.
[55] D C Gibbons, *The Criminological Enterprise: Theories and Perspectives* (1979).
[56] E F Greenberg, "On one-dimensional Marxist criminology" (1976) 3 Theory and Society 610 at 612.
[57] K J Hayward and J Young, "Cultural criminology: some notes on the script" (2004) 8 Special Edition of the International Journal Theoretical Criminology 259.
[58] M R Maidment, "Feminist Perspectives in Criminology", in W S DeKeseredy and B Perry (eds), *Advancing Critical Criminology: Theory and Application* (2006) 43.
[59] J Lea and J Young, *What is to be Done About Law and Order: Crisis in the Eighties* (1984) 262.
[60] R Quinney, "Crime Control in Capitalist Society: A Critical Philosophy of Legal Order", in I Taylor, P Walton and J Young (eds), *Critical Criminology* (1975) at 54.

underclass are not only linked with the deprivation of the perpetrator. There also exist many crimes that can be attributed to a sense of envy.[61] By suggesting the commission of crimes as "reproduction of the capitalist system",[62] critical criminology arguably over-identifies with the underdog at the cost of ignoring the reality. Additionally, this interconnectedness between crimes and social inequality does not offer any answer as to why social elites commit crimes.[63]

Furthermore, critical criminology also misses the fact that law-breakers are essentially doing wrong. By concentrating on the unfairness and injustice experienced by offenders, critical criminologists have seemingly ignored the real victims and their innocent sufferings,[64] such as the innocent pedestrian involved in a mugging, or shop-owners robbed unguardedly. Moreover, not all laws are inherently oppressive. Some legislation is designed to protect citizens and promote justice. Critical criminology merely has "translated wide varieties of social, legal and cultural conflict into class conflict, and strained to find class conflict in situations of overwhelming consensus".[65] By focusing the difference on political consciousness, further ambiguity might be caused without a thorough explanation for the cause of crime.

(3) Ineffectiveness and idealism

The current marginal status of critical criminology can be traced back to its inherent problem, namely, its idealism.[66] By pinning the cause of crime on the inequality between two classes, critical criminologists seem to suggest that by changing a society to one where everyone shares everything, the problem will fade away.[67] As argued, "socialist societies should have much lower rates of crime because the less intense class struggle should reduce the forces leading to and the functions of crime".[68] Schwendinger and Schwendinger also shared the same view, stating that "to resolve the contradictions and the expression in rape of the social psychological effect of these contradictions, is to change a class system of oppression in America today to a socialist political economy and a relatively crime-free tomorrow".[69] The insistence on finding an interconnectedness between crime and capitalist society, however, is criticised as a determinism "shifted from the socio-cultural context to the realm of political economy".[70] By claiming that the removal of economic exploitation will eliminate the main cause of crime, critical criminology can be described as reductionist and over-simplistic. Firstly, as mentioned, it neglects other forces in society that might contribute to crime and secondly, in reality, crimes will not stop occurring simply because of a society's changing nature.[71] For example, the crime rate in Sweden rose shortly after a social-democratic system replaced its capitalist one.[72]

This utopian hope is further crushed by some criminologists who argue that any reduction in crime in socialist states is the result of sternly repressive enforcement, rather than an attempt to reduce

---

[61] J Toby, "The New Criminology is the Old Baloney", in J A Inciardi (ed), *Radical Criminology: The Coming Crises* (1980) 124.
[62] Quinney (n 60) at 54.
[63] R L Akers, "Further Critical Thoughts on Marxist Criminology: Comments on Turk, Toby, and Klockars", in J A Inciardi (ed), *Radical Criminology: The Coming Crises* (1980).
[64] Lea & Young, *What is to be Done* (n 59).
[65] Klockars (n 52) at 113.
[66] M J Lynch and W B Groves, *A Primer in Radical Criminology* (1986).
[67] Taylor et al, The New Criminology (n 12).
[68] Chambliss (n 32) at 153.
[69] J R Schwendinger and H Schwendinger, "Rape myths: in legal, theoretical, and everyday practice" (1974) 1 Crime and Social Justice 18 at 25.
[70] Dwones & Rock, *Understanding Deviance* (n 38) 244.
[71] D Shichor, "Some Problems of Credibility in Radical Criminology", in J A Inciardi (ed), *Radical Criminology: The Coming Crises* (1980) 191 at 196.
[72] R L Akers and C S Sellers, *Criminological Theories: Introduction, Evaluation, and Application*, 5th edn (2009).

conflict.[73] In other words, the conflicts within society can still be witnessed in socialist countries. Critical criminology, thus, cannot explain the criminality that exists in states where private ownership of the means of production has been abolished.[74] The panacea that critical criminologists imagined is in fact ineffective. It is arguably useless not only because it, in reality, does not work, but also because it is a massive move. Recommending the total erasure of a current social structure and its replacement with socialism is not a solution that can effortlessly be put into practice. The theory is inevitably reprimanded as ineffective and impractical as it fails to provide any short-term, immediate resolution to solve the problem of crime.[75]

The marginal status of critical criminology is the result of its inability to respond to social, political and economic transformations around the globe and the corresponding changes in crime control,[76] such as the rise of private policing, evidence of "active citizenship", issues regarding "minority groups" and so forth.[77] The failure to progress to a new political agenda and the laborious attention on crime control *per se* leads critical criminology to become more of an analytical practice than a forwarding theory which can be utilised in current society.[78]

E. MARGINAL AS AN APPROACH TO CRIMINOLOGY; INEVITABLE AS A PERSPECTIVE

Critical criminology is undeniably under attack from different perspectives, as illustrated. Some of its defects are arguably responsible for its current marginal status in the field of criminology. That said, the essence of critical criminology is to analyse disparate concepts from a conflicting point of view and to be able to reject any previous thoughts This is still being frequently employed. For example, the spirit of cultural criminology, which is based on engaging in critical debates with policies and other approaches to criminology, can be positively traced back to the perspective advanced in critical criminology.[79] The concept of inequality in society, which is one of the key advancements in critical criminology, is also in the heart of feminist movements. Feminist criminology was indeed focusing on the inequality of sex within a society and critically reprimanding other theories for ignoring such a disadvantage. This article concludes that the essence of critical criminology is still virtually insightful and at the heart of generating more effective and convincing criminology, albeit its lack of relevance *per se*.

F. CONCLUSION

Crime, as a frequently encountered societal phenomenon, deserves to be explained and analysed. Different criminologists have offered valuable insights regarding the causes of crime. Before the birth of critical criminology, mainstream criminologists attributed the causes of crime to individual factors. It was considered that crimes are committed by individuals who are either influenced by different factors or rationally choose to do so after weighing the outcomes against the inputs. Critical criminology, as the name suggests, took a critical stand against all mainstream approaches to criminology. According to critical criminologists, crime should be examined in the context of a

[73] S Cohen, "The punitive city: notes on the dispersal of social control" (1979) 3 Contemporary Crisis 339.
[74] *Ibid.*
[75] Lea & Young, *What is to be Done* (n 59).
[76] D Garland and R Sparks, "Criminology, social theory and the challenge of our times" (2000) 40 The British Journal of Criminology 189.
[77] D Garland and R Sparks, "Introduction", in D Garland and R Sparks (eds), *Criminology and Social Theory* (2001) at 191-201.
[78] R Hill and R Robertson, "What sort of future for critical criminology?" (2003) 39 Crime, Law & Social Change 91.
[79] J Ferrell, "Cultural criminology and the politics of meaning" (2013) 21(3) Critical Criminology 257.

political and economic capitalist society. Instead of exploring the cause of crime starting from individuals, critical criminology believes it can readily be identified from class conflicts and struggles within the capitalist society. The role of individuals, from a critical criminologist's perspectives, can be minimised and even replaced by a more comprehensive construction of society as a whole. Critical criminology has offered valuable insights into the phenomenon through a shift of concentration, asserting that the initial criminalisation of particular behaviours should replace the focus of criminologists on the causes of crime. While denying the meaning of deviance as abnormal behaviours, critical criminology advanced that the concept is indeed a social construction. Adopting some Marxist perspectives, it argued that deviance is effectively a concept created by groups of economically and politically influential people to maintain their entitled life.

Whilst being insightful, critical criminology is not immune to criticisms and has been attacked from different aspects. Firstly, its failure to maintain its contentions has been identified. The possibly plausible definition conferred to deviance is also criticised as being tautological and unsatisfactory. It is additionally suggested that the cause of crime advanced by critical criminology cannot satisfactorily explain reality, especially in a world of globalisation and growing international consensus. Lastly, critical criminology is also criticised as being unable to provide any short-term solutions to the problems of crime. A complete societal revolution is argued to be impractical and useless, entailing the ineffectiveness of critical criminology. Despite the theory's marginal status, it is arguably still at the heart of criminology.

TRIAL OF TRUTH: LAW AND 'FAKE NEWS'

*Chloe Francis**

A. INTRODUCTION

The threat of 'fake news' has elicited widespread fears about its effect on our society, our politics, and our national security. Claims of 'fake news' dominated recent national elections around the world, notably the 2016 United States (US) Presidential Election,[2] the 2017 British General Election,[3] and the 2017 French Presidential Election.[4] Worries of 'fake news' unite people worldwide— a BBC World Service poll found that 79% of respondents from eighteen surveyed countries worry about what is real and what is fake on the internet.[5] Additionally, 46% of news audiences across the United Kingdom (UK), the US, France, and Brazil believe that 'fake news' had at least "somewhat" influenced their national elections.[6] There is an identified fear and a perceived threat of 'fake news'. As such, policymakers around the world endeavour to find a solution to 'fake news'.[7] However, these

* 4th year LLB student at the University of Edinburgh.
[2] H Allcott and M Gentzkow, "Social media and fake news in the 2016 election" (2017) 31(2) J of Economic Perspectives 211.
[3] L Byrne, "A post-truth era? Fake news and the general election" (16 May 2017) ICM Unlimited, available at *https://www.icmunlimited.com/polls/post-truth-era-fake-news-general-election/*.
[4] P N Howard et al, "Junk news and bots during the French presidential election: what are French voters sharing over twitter in round two?' (2017) Oxford Internet Institute, available at *http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/05/What-Are-French-Voters-Sharing-Over-Twitter-Between-the-Two-Rounds-v9.pdf*; Bakamo, "The Role and Impact of Non-Traditional Publishers in the French Elections 2017" (2017), available at *https://static1.squarespace.com/static/58495e3329687f8bfbb3f25c/t/58f5b4cd2994ca075dfa803c/1492497618893/Role+and+Impact+of+Non-Traditional+Publishers+in+the+French+Presidential+Election+-+Report+1+-+Bakamo.pdf.*
[5] R Cellan-Jones, "Fake news worries 'are growing' suggests BBC poll" (22 September 2017) BBC News, available at *http://www.bbc.co.uk/news/technology-41319683*.
[6] Kantar, "Trust in News" (2017) at 35, available at http://www2.kantar.com/l/208642/2017-10-27/6g28j.
[7] Culture, Media and Sport Committee, "Fake News Inquiry Launched" (2017a), available at *https://www.parliament.uk/business/committees/committees-a-z/commons-select/culture-media-and-sport-*

attempts are plagued with difficult questions: what is 'fake news', and what problem does it pose? This article will not only probe these questions but will also use that investigation as a basis to explore what the law can do to address the problem of 'fake news' through direct and indirect regulation. The author relies on Lawrence Lessig's theory of regulation that posits that law is one of four interdependent modalities of regulation, the others being the market, social norms, and code, i.e. physical or technical restraints.[8] Independent or collective pressure from these modalities can dictate behaviour. Part B of the article will explore the definition of 'fake news' and the identification of the problem it presents. Part C will analyse the ability of the law to directly address 'fake news' through consideration of freedom of expression. Part D will also explore direct regulation outside of human rights concerns. Finally, Part E will consider what the law can do indirectly. This article has particular focus on British law, and its analysis suggests that, although there are many ways in which the law can address 'fake news', each method is limited by considerations of effectiveness and democratic costs.

It should be noted that, to date, the most explicit and authoritative guidance on how to approach 'fake news' is the Joint Declaration on Freedom of Expression and "Fake News", Disinformation and Propaganda (Joint Declaration) issued by the UN Special Rapporteur on Freedom of Opinion and Expression and other regional human rights experts in March 2017.[9] Although Joint Declarations are not legally binding, they are highly persuasive[10] interpretations of freedom of expression and have been relied on by the European Court of Human Rights (ECtHR) in the past.[11] Therefore, discussion of any direct legal action will consider the Joint Declaration.

B. THE PROBLEM OF 'FAKE NEWS'

The problem of 'fake news' must be identified to define the scope of any response, direct or indirect. This requires consideration of what 'fake news' is and an assessment of the threat that it poses.

---

committee/news-parliament-2015/fake-news-launch-16-17/; Culture, Media and Sport Committee, "Fake News" (2017b), available at *https://www.parliament.uk/business/committees/committees-a-z/commons-select/digital-culture-media-and-sport-committee/inquiries/parliament-2017/fake-news-17-19/*; European Commission, "Public Consultation on Fake News and Online Disinformation" (2017), available at *https://ec.europa.eu/info/consultations/public-consultation-fake-news-and-online-disinformation_en*; Ministry of the Interior of the Czech Republic's Centre Against Terrorism and Hybrid Threats, "Terorismus a Měkké Cíle" (2017), available at *http://www.mvcr.cz/cthh/clanek/centre-against-terrorism-and-hybrid-threats.aspx*; S B Yi, "New legislation to combat fake news likely to be introduced next year: Shanmugam" (19 June 2017) The Straits Times, available at *http://www.straitstimes.com/singapore/new-legislation-to-combat-fake-news-next-year-shanmugam*; C Edwards, "Italy debates fines and prison terms for people who spread fake news" (16 February 2017) The Local, available at *https://www.thelocal.it/20170216/italy-mulls-introducing-fake-news-fines*; A Chrisafis, "Emmanuel Macron promises ban on fake news during elections" (3 January 2018) The Guardian, available at *https://www.theguardian.com/world/2018/jan/03/emmanuel-macron-ban-fake-news-french-president*.

[8] L Lessig, "The law of the horse: what cyberlaw might teach" (1999) 113(2) Harv L Rev 501; L Lessig, *Code: Version 2.0* (2006).

[9] Organization for Security and Co-operation in Europe, Joint Declaration on Freedom of Expression and 'Fake News', Disinformation and Propaganda (2017), available at *http://www.osce.org/fom/302796?download=true*.

[10] T Mendel, "The UN Special Rapporteur on Freedom of Opinion and Expression: Progressive Development of International Standards Relating to Freedom of Expression", in T McGonagle and Y Donders (eds), *The United Nations and Freedom of Expression and Information: Critical Perspectives* (2015).

[11] *Youth Initiative for Human Rights v Serbia* (2013) Application no. 48135/06; *Editorial Board of Pravoye Delo and Shtekel v Ukraine* (2011) Application no. 33014/05; *Stoll v Switzerland* (2007) Application no. 69698/01; *Yildirim v Turkey* (2012) Application no. 3111/10; Mendel (n 10).

(1) What is 'fake news'?

There is no universal or consistent definition of 'fake news', despite the public's general awareness. Both the UK[12] and the European Union (EU)[13] public consultations regarding 'fake news' raise the question of definition. Moreover, the Joint Declaration is silent on the matter and, therefore, cannot function as a point of guidance. There is a growing body of literature from academic institutions and media non-profit organisations that indicates that 'fake news' can include, but is not limited to, complete fabrications,[14] misleading content,[15] satire or parody,[16] foreign interference in domestic politics,[17] and bad journalism.[18] However, even a taxonomical approach contributes very little to understanding or defining 'fake news' because, in reality, the lines between the different types blur together. An article may include both misleading content and satire, or a mixture of fabrication and misrepresentation. Nevertheless, despite the many forms of 'fake news', the harm caused is universal: the deception of the public. On that basis, a broad-brush approach could be argued; however, that would fail to acknowledge the varying levels of seriousness and risk punishing legitimate forms of speech. For example, satire and parody, powerful tools of social and political commentary, should not be treated similarly to complete fabrications. Moreover, one could look to the underlying motivations behind 'fake news', rather than characteristics of form, for the basis of a definition. However, although there are identifiable financial, political, or mischievous motivations behind 'fake news', they are blurred together as well. Therefore, this conflation frustrates attempts at definition on the basis of motivation. The breadth of 'fake news', its interwoven nature, and its multiple signifiers are why the term is so hard to define. It is nonetheless necessary for the law to clearly and narrowly define 'fake news' to prevent overregulation that could chill speech and to comply with the rule of law.[19]

At its core, 'fake news' is misinformation, which is defined as "false or inaccurate information, especially that which is deliberately intended to deceive".[20] Analysis of various definitions of 'fake news', including that of the Ethical Journalism Network[21] and scholarly articles,[22] reveals deceit to be a central element of 'fake news'. This article, though, will make a distinction between an intent to deceive and the effect of deceit and will broadly define 'fake news' as 'deceitful misinformation disseminated to the public, relating to news or current events'.[23] Although the author is of the view that an intent to deceive is a necessary component of 'fake news' and that any legal

[12] Culture, Media and Sport Committee, *Fake News* (2017a) (n 7); Culture, Media and Sport Committee, *Fake News* (2017b) (n 7).
[13] European Commission, *Public Consultation* (n 7).
[14] C Wardle, "Fake news. It's complicated" (16 February 2017) Medium, available at *https://medium.com/1st-draft/fake-news-its-complicated-d0f773766c79*; EAVI, "Infographic: Beyond 'Fake News' - 10 Types of Misleading News" (31 October 2017), available at *https://eavi.eu/wp-content/uploads/2017/07/beyond-fake-news_COLOUR_WEB.pdf*.
[15] Wardle (n 14); EAVI (n 14); D Lilleker, "Evidence to the Culture, Media and Sport Committee 'fake news' inquiry" (2017) *Centre for Politics & Media Research, Faculty for Media & Communication, Bournemouth University*, available at *http://eprints.bournemouth.ac.uk/28610/3/Evidence%20Submission%20-%20Fake%20News%20FINAL.pdf*.
[16] D Tambini, "Fake news: public policy responses" (2017) The London School of Economics and Political Science Media Policy Brief, available at *http://eprints.lse.ac.uk/73015/1/LSE%20MPP%20Policy%20Brief%2020%20-%20Fake%20news_final.pdf*; Wardle (n 14); M Verstraete, D E Bambauer and J R Bambauer, "Identifying and countering fake news" (2017) The University of Arizona, available at *https://law.arizona.edu/sites/default/files/asset/document/fakenewsfinal_0.pdf*; EAVI (n 14); Lilleker (n 15).
[17] Tambini (n 16).
[18] *Ibid*; Wardle (n 14).
[19] Organization for Security and Co-operation in Europe, *Joint Declaration* (n 9).
[20] Definition of "misinformation", available at *https://en.oxforddictionaries.com/definition/misinformation*.
[21] Ethical Journalism Network, 'Fake news': Information Disorder – Disinformation - Misinformation - Mal-information, available at *http://ethicaljournalismnetwork.org/tag/fake-news.*
[22] S Burshtein, "The true story on fake news" (2017) 29 IPJ 397; Allcott & Genztkow (n 2).
[23] The focus on news and current events will be explained below.

definition should include such an intent as it narrows the scope to culpable instances of 'fake news'; for the purposes of this article an intent to deceive is excluded from the working definition of 'fake news'. This is because specifying an intent to deceive could exclude bad journalism, as these failings can also result from error or bad practice. It is important that bad journalism is included within the scope of this academic discussion because modern journalistic practices are contributors to the problem; this is reflected in the public's perception that traditional news media plays an active role in spreading 'fake news'.[24]

(2) What is the problem?

There is a dearth of evidence with respect to the prevalence, spread, and effect of 'fake news' in the UK and Europe. This is partly because of both the novelty of the issue and that data is being withheld[25] by relevant online platforms. Due to the fact that the majority of 'fake news' is disseminated online,[26] this data is key to understanding the scale and effect of 'fake news'. Most available research focuses on the US landscape, with regard to the 2016 US Presidential Election.[27] More evidence is needed about 'fake news' in the UK before it can be addressed directly by law. Nevertheless, there is a perceived threat of harm that has elicited public and governmental discomfort. That threat needs to be identified to provide a basis for legal intervention.

The unrealised problem of 'fake news' is the deteriorating effect it has on democracy.[28] The foundation of democracy is public choice enabled by free and unhindered debate. At a fundamental level, 'fake news' subverts the autonomy of individuals, an important aspect of civil liberty and the basis of democratic choice. Joseph Raz posits that autonomy derives from both the availability of choice and that choice being "free from coercion and manipulation by others".[29] 'Fake news' causes the public to involuntarily consume and believe information that they would not otherwise. Although it is well-known that the public self-selects news that conforms to their pre-existing views,[30] their selection of biased reporting is nonetheless based on an assumption that the content is founded on truth, not falsity. Therefore, the choice of these individuals is not the product of free, independent thought; it is the product of unwitting manipulation. Moreover, although the effect 'fake news' has on autonomy is individual, it bears a potential cumulative effect on democracy as a whole. 'Fake news' undermines democracy on an interpersonal level. By devaluing and delegitimising the voices of "expertise, authoritative institutions, and the concept of objective data",[31] 'fake news' frustrates "society's ability to engage in rational discourse",[32] which is an instrumental feature of democracy.[33] In sum, what is clear is that there is an identifiable threat of harm; however, even if proven, the intangibility of that harm presents problems in providing adequate justification for legal intervention.

---

[24] Kantar, *Trust* (n 6) 33.
[25] J Schwartz, "Facebook undermines its own effort to fight fake news" (7 September 2017) Politico, available at *https://www.politico.com/story/2017/09/07/facebook-fake-news-social-media-242407.*
[26] A Guess, B Nyhan and J Reifler, "Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 U.S. presidential campaign' (2018) European Research Council, available at *http://www.dartmouth.edu/~nyhan/fake-news-2016.pdf*.
[27] A Schiffrin, "How Europe fights fake news" (26 October 2017) Columbia Journalism Review, available at *https://www.cjr.org/watchdog/europe-fights-fake-news-facebook-twitter-google.php*.
[28] A Chrisafis, "Emmanuel Macron promises ban on fake news during elections' (3 January 2018) The Guardian, available at *https://www.theguardian.com/world/2018/jan/03/emmanuel-macron-ban-fake-news-french-president*.
[29] J Raz, *The Morality of Freedom* (1988) 373.
[30] J Hsu, "People choose news that fits their views" (7 June 2009) Live Science, available at *https://www.livescience.com/3640-people-choose-news-fits-views.html*; P E Converse, "The nature of belief systems in mass publics (1964) 18 Critical Review 1.
[31] S Baron and R Crootof, "Fighting fake news" (2017) Information Society Project at Yale Law School and the Floyd Abrams Institute for Freedom of Expression Workshop Report at 3, available at *https://law.yale.edu/system/files/area/center/isp/documents/fighting_fake_news_-_workshop_report.pdf*.
[32] *Ibid*.
[33] *Ibid.*

The threat 'fake news' poses toward democracy is grounded in a specific type of misinformation: that relating to news or current events. Accordingly, any legal intervention should be narrowly constructed. False information about specific individuals is covered by defamation law and other types of misinformation do not pose the same type of threat as 'fake news'. Attempts to restrict benign misinformation would be unjustified and contrary to freedom of expression.

C. FREEDOM OF EXPRESSION

Freedom of expression is a fundamental human right and a cornerstone of democracy. As any action against 'fake news' would constitute a restriction of speech and the spread of information, freedom of expression is engaged. It is the primary concern for any legal intervention toward speech. Direct legal action is dependent on a basis for exception within freedom of expression. The freedom is particularly restrictive due to the binding international framework. Article 10 of the European Convention on Human Rights (ECHR) confers freedom of expression as a freedom "to hold opinions and to receive and impart information and ideas without interference".[34] The ECHR is given direct effect in the UK through implementation by the Human Rights Act 1998[35] and is enforceable through the ECtHR. Moreover, the freedom remains an important consideration where 'fake news' is indirectly addressed because even actions of the private sector are subject to intense scrutiny where restrictions of speech are involved.

Article 10 does not contain any qualification as to truth, and the ECtHR has yet to adjudicate directly on whether misinformation or 'fake news' is protected under Article 10. There is a case[36] pending on whether 'fake news' bans are compatible with Article 10 but the ECtHR is yet to make a judgment. Moreover, *Salov v Ukraine*[37] tangentially addresses 'fake news'. The case concerned a citizen's discussion of false information contained in a newspaper which was not produced or published by him. The ECtHR held that Article 10 does not preclude the dissemination of already received information of dubious veracity.[38] The case is relevant to the public's sharing of 'fake news' — particularly on social media platforms — but has no bearing on the position of those who create and first distribute 'fake news'. This article will next explore the degree to which the ECHR protects 'fake news' and will argue that freedom of expression can preclude 'fake news'.

(1) Freedom of expression and 'fake news'

It is necessary to examine whether freedom of expression is conceptually incompatible with 'fake news' as a preliminary matter because any such incompatibility serves as a basis for subsequent arguments. Freedom of expression operates on two levels. It enables the development[39] and self-fulfilment[40] of individuals by guaranteeing access to information and ideas, and through ensuring an educated, free-thinking populace, freedom of expression forms the foundation of the maintenance and progression of democracy. Freedom of expression's operation reveals its purpose of enabling and promoting principles that are essential for democracy: freedom of political debate[41] and "pluralism, tolerance, and broadmindedness".[42] The proper enjoyment of the freedom is implicitly dependent on the truthful nature of information spread; only on a shared foundation of truth and mutually agreed upon facts can citizens be adequately informed, intellectually engaged, and respectfully tolerant of the differences of others. The alleged prevalence of 'fake news' — or even the fear of 'fake news' —

[34] European Convention for the Protection of Human Rights and Fundamental Freedoms, Art 10
[35] Human Rights Act 1998, s 1(1)(a).
[36] *Dareskizb Ltd v Armenia* (2008) Application no. 61737/08.
[37] *Salov v Ukraine* (2005) Application no. 65518/01.
[38] *Ibid*. at para 113.
[39] *Handyside v The United Kingdom* (1976) Application no. 5493/72 at para 49.
[40] *Lingens v Austria* (1986) Application no. 9815/82 at para 41.
[41] *Ibid*. at para 42.
[42] *Handyside* (n 39) at para 49.

prevents trust in facts presented by alternative sources, i.e. sources that do not align with previously held views. Therefore, 'fake news' corrupts the shared foundation and undermines the core values and purpose of freedom of expression.

(2) Limitations on freedom of expression

The incompatibility of 'fake news' with freedom of expression is only justification for legal intervention in principle. In practice, there is a set framework and a binding body of case law for its limitation.

(a) Article 10(2)

Article 10(2) of the ECHR imposes three cumulative conditions that any lawful condition, restriction, or penalty on freedom of expression must satisfy.[43] The interference must be prescribed by law, aimed at protecting one of the specified legitimate interests, and necessary in a democratic society. The first and second conditions will be considered in detail. The above discussion on the incompatibility of 'fake news' with democratic principles sufficiently explores whether legal interference is 'necessary in a democratic society'. Therefore, further discussion is unnecessary.

Whether interference is 'prescribed by law' is a matter of construction and clarity. Laws must be "sufficiently precise to enable citizens to reasonably foresee the consequences".[44] Therefore, the definitional difficulties, as discussed in Part B, are formally inhibitive of the law's ability to directly address 'fake news'. The Joint Declaration makes clear that the definition of 'fake news' must extend past its quality of falseness.[45] Falseness is inherently too vague,[46] as it is a fluid concept that can change over time depending on the facts available. Clear definition of 'fake news' must, therefore, rely on other qualities, such as form, to narrow its scope. This has been exhibited in the proposed Italian 'fake news' bill, which focused on news reports and excluded the print media; however, the bill failed to provide requirements of intent or knowledge, which left the definition overly broad and led to criticism.[47] As argued above, an intent to deceive is necessarily narrow and would limit 'fake news' laws to culpable instances. Although difficult, a clear definition of 'fake news' is not impossible; it merely requires a high level of detail. As breadth frustrates clarity, a clear definition could be unable to capture every type of 'fake news'. Nevertheless, total regulation is unnecessary for effective regulation. Effective regulation can be achieved through the reduction of undesirable behaviour below a desirable threshold. Although 'fake news' is hard to define, a clear definition is possible; therefore, the 'prescribed by law' condition is not totally inhibitive of the law's ability to address 'fake news' directly.

In regard to the second condition of Article 10(2), although the legitimate aims focus on either the public interest or the rights of others, none are sufficiently broad to address 'fake news'. Nonetheless, there are established limitations that incidentally engage with it. Defamation laws and hate speech restrictions, which pursue the legitimate aim of protecting the "reputation or rights of others",[48] can serve to deny protection to 'fake news' that includes defamatory or hate speech. As deception does not engage with any legally recognised rights of others, 'fake news' devoid of defamatory or hate speech could not fall into this legitimate aim. Additionally, 'fake news' deliberately created by foreign powers to interfere with domestic elections could warrant legal intervention on the basis of the legitimate aim of protecting national security. Claims of foreign

[43] Art 10(2) of the European Convention for the Protection of Human Rights and Fundamental Freedoms.
[44] S Greer, *The Exceptions to Articles 8 to 11 of the European Convention on Human Rights* (1997) 11.
[45] Organization for Security and Co-operation in Europe, *Joint declaration* (n 9) at para 2(a).
[46] *Ibid*.
[47] Edwards (n 7).
[48] Art 10(2) of the European Convention for the Protection of Human Rights and Fundamental Freedoms.

intervention are still being investigated,[49] but there is an already-identified threat of foreign powers deliberately attempting to spread misinformation.[50] However, it is doubtful that, even if proven, foreign interference through 'fake news' would fall within the scope of the national security interest. The Johannesburg Principles on National Security, Freedom of Expression and Access to Information, a persuasive interpretive tool of the ECHR that has been approved by the United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression, state that restrictions on freedom of expression are only justified "if a government can demonstrate that…the expression is intended to incite imminent violence".[51] Although there have been instances where 'fake news' has shown the potential to incite violence,[52] such events are not only anecdotal but were unforeseen and without the requisite intent. Restrictions on 'fake news' under the 'national security' legitimate aim would only come about in special circumstances.

In sum, the limitations under Article 10(2) have very narrow application and do not have the capacity to confront 'fake news'. Although Article 10(2) fails to provide a basis for exception, it is nonetheless revealing of the nature of freedom of expression. By framing the limitations through the pursuit of legitimate democratic aims, Article 10(2) reveals the consequentialism behind this freedom: expression is only free if it serves a democratic purpose. This consequentialist approach will form the bedrock for the exception of 'fake news', which will be explored below.

(b) Obligations of journalists

Although Article 10 does not explicitly include freedom of the press, the ECtHR has granted the press a special status through the development of case law.[53] This is due to the integral role the press plays as a vehicle for communication and, therefore, freedom of expression — at least before the internet. Especially in imparting current and relevant political information, the press serves a crucial democratic role in enabling freedom of debate within the populace. As such, the ECtHR has conferred a duty onto the press "to impart information and ideas on matters of public interest",[54] which gives the press greater freedom by limiting the possibilities for states to restrict it.[55] However, that greater freedom is subject to the "duties and responsibilities"[56] attached to the exercise of freedom of expression; although not elaborated on in the ECHR, these "duties" are limitations to freedom of expression that are wholly dependent on interpretation and context. These "duties" as they relate to the press require journalists to act "in good faith and on an accurate factual basis" to provide "'reliable and precise information in accordance with the ethics of journalism".[57] Therefore, journalists are

[49] D Sabbagh, "Facebook to expand inquiry into Russian influence of Brexit" (17 January 2018) The Guardian, available at *https://www.theguardian.com/technology/2018/jan/17/facebook-inquiry-russia-influence-brexit*.
[50] The EU has established a disinformation review in a campaign to confront Kremlin misinformation (EU vs Disinformation), available at *https://euvsdisinfo.eu/*; O Solon and S Siddiqui, "Russia-backed Facebook posts 'reached 126m Americans' during US election" (21 October 2017) The Guardian, available at *https://www.theguardian.com/technology/2017/oct/30/facebook-russia-fake-accounts-126-million*.
[51] Art 19 of the Global Campaign for Free Expression, "The Johannesburg Principles on National Security, Freedom of Expression and Access to Information" (1996) at 8, available at *https://www.article19.org/data/files/pdfs/standards/joburgprinciples.pdf*.
[52] E Lipton, "Man motivated by 'pizzagate' conspiracy theory arrested in Washington gunfire" (5 December 2016) The New York Times, available at *https://www.nytimes.com/2016/12/05/us/pizzagate-comet-ping-pong-edgar-maddison-welch.html*; R Goldman, "Reading fake news, Pakistani minister directs nuclear threat at Israel' (24 December 2016) The New York Times, available at *https://www.nytimes.com/2016/12/24/world/asia/pakistan-israel-khawaja-asif-fake-news-nuclear.html*.
[53] *Lingens v Austria* (1986) Application no. 9815/82; *Castells v Spain* (1992) Application no. 11798/85; *Thorgeir Thorgeirson v Iceland* (1992) Application no. 13778/88; D Bychawska-Siniarska, "Protecting the right to freedom of expression under the European Convention on Human Rights" (2017) Council of Europe at 87, available at *https://rm.coe.int/handbook-freedom-of-expression-eng/1680732814*.
[54] *Observer and Guardian v The United Kingdom* (1991) Application no. 13585/88 at para 59.
[55] Bychawska-Siniarska (n 53) at 98.
[56] Art 10(2) of the European Convention for the Protection of Human Rights and Fundamental Freedoms.
[57] *Fressoz and Roire v France* (1999) Application no. 29183/95 at para 54; see further *Bergens Tidende and Others v Norway* (2000) Application no. 26132/95 at para 53.

precluded from purposefully publishing misinformation. These obligations of journalists are projections of freedom of expression's consequentialism. Due to the power they hold, journalists only enjoy freedom of expression as long as that expression continues to support the freedom's democratic objectives. Ultimately, these "duties" provide an apt basis for legal restriction of 'fake news' within traditional media; however, the internet has changed the operation of media and contributed significantly to the problem of 'fake news'. Online, the obligations of journalists are of limited value as individuals are increasingly adopting the role of traditional media and disseminating information themselves.

The ECtHR has explicitly stated that Article 10 applies to online communication[58] and has stressed that, due to the increased scale of information individuals encounter online, the obligations of journalists take on "added importance".[59] However, the ECtHR has failed to acknowledge the internet's effect on journalism itself. In the online environment, journalism has been democratised. The wide availability of free platforms on which users can publish information and ideas has enabled anyone to be a 'journalist' operating outside the prescribed standards and ethics of professional journalism. However, although there has never been formal legal definition of what constitutes 'journalism' by either the UK or the ECtHR, to categorise everyone with a keyboard as a 'journalist' would be an obvious stretch of the term. The Oxford Dictionary of English defines 'journalist' as "a person who writes for newspapers, magazines, or news websites or prepares news to be broadcast".[60] As such, it is clear that the notion of 'journalist' is closely intertwined with traditional media institutions. Therefore, in that respect, journalists can be distinguished from other producers of information online whom are not associated with traditional media institutions, like bloggers. As a distinct and influential class of people in this new media environment, internet users should be subject to "duties and responsibilities" in exercise of freedom of expression — even though the ECtHR has not recognised those "duties" as they relate to individuals. The public nature and viral potential of online content places internet users and professional journalists on an equal footing in terms of ability, reach, and influence; therefore, users who report news and current events should have obligations akin to journalists relating to the accuracy and reliability of the information they spread. It is acknowledged that, unlike professional journalists, individuals do not have general duties that limit freedom of expression. This is the main basis of the argument against the imposition of "duties" on internet users. However, journalistic duties are nothing more than institutionalised reiterations of the "duties and responsibilities" attached to the freedom of expression. The imposition of accuracy and reliability obligations on internet users who report on matters of public importance would be a natural extension of the consequentialist approach inherent in freedom of expression as it would ultimately promote democratic principles. As such, states are justified in conferring "duties and responsibilities" to internet users. The freedom of states to develop this new limitation on freedom of expression is the basis of future legal action against 'fake news' online.

(3) Article 17

Rather than operating within the limits of Article 10, it could be argued that 'fake news' could fall outside the scope of freedom of expression entirely. Article 17 of the ECHR excludes any activity or act "aimed at the destruction of any of the rights and freedoms" of the ECHR from protection as *prima facie* inadmissible.[61] Although strict construction of Article 17 suggests that the action would have to relate to a specific right or freedom, the ECtHR has made clear that Article 17 applies to any actions that are "incompatible with the values proclaimed and guaranteed by the Convention".[62] As has been argued, 'fake news' undermines both freedom of expression and the democratic principles it promotes; however, it is unlikely that it is destructive in the way recognised by Article 17. 'Fake

[58] *Ashby Donald and Others v France* (2013) Application no. 26132/95 at para 34.
[59] *Stoll v Switzerland* (2007) Application no. 69698/01.
[60] Definition of "journalist", available at https://en.oxforddictionaries.com/definition/journalist.
[61] European Convention for the Protection of Human Rights and Fundamental Freedoms, Art 17; *Seurot v France* (2004) Application no. 57383/00.
[62] *Pavel Ivanov v Russia* (2007) Application no. 35222/04 at para 1.

news' is fundamentally different from the threats already excluded by Article 17, such as racial hate speech[63] or historical negationism and revisionism,[64] because it does not explicitly espouse ideas contrary to the ECHR. 'Fake news' is simply an instrument of disruption. Therefore, it may not be a sufficiently obvious threat to democracy to warrant expulsion under Article 17. Moreover, one could argue that 'fake news' could fall within established excluded speech. Historical negationism and revisionism are forms of misinformation and could provide space for 'fake news' within the ambit of Article 17. However, the ECtHR has delineated the limits of historical negationism and revisionism as only applying to a "category of clearly established historical facts".[65] As the bulk of 'fake news' focuses on current events, only a small sector of 'fake news' is likely to fall under the "category of clearly established historical facts".

In conclusion, generally freedom of expression only narrowly precludes 'fake news' where it amounts to defamation and hate speech, but there is broader protection where the obligations of journalists are engaged. Especially online, the uneven boundaries of freedom of expression only provide piecemeal protection against 'fake news' and are, therefore, inhibitive of generalised restrictions. However, it has been argued that there is space for further development of the limitations of the freedom with respect to online users because of their heightened and unprecedented influence in cyberspace. One could argue that the positive obligation of states "to create [a] favourable environment"[66] for the exercise of freedom of expression obliges states to develop and impose these obligations. Therefore, with the view that states have the ability to impose online user obligations, the law can directly address 'fake news' without infringing upon freedom of expression. Moreover, any legal intervention must include a clear definition of 'fake news' so that it is prescribed by law. The above considerations, especially the fundamental incompatibility of 'fake news' with the freedom of expression, provide adequate justification for indirect action as well.

D. CAN LAW DIRECTLY ADDRESS 'FAKE NEWS'?

It has been established that the law can address 'fake news' within the ambits of freedom of expression. This Part will further explore what the law can do directly and with a special consideration of how it can address 'fake news' within traditional media.

(1) 'Fake news' laws

Although the law can directly address 'fake news' under freedom of expression, this approach has limitations. Notably, 'fake news' laws cannot oblige online platforms to actively monitor content posted or shared within them. General obligations to monitor are precluded under Article 15 of the E-Commerce Directive.[67] Additionally, the law cannot prohibit the dissemination of information that has already been received.[68] Therefore, 'fake news' laws must only target the creation and initial dissemination of content.

Although the law can directly address 'fake news', it cannot do so effectively because the law is unable to regulate cyberspace. The scale of information online greatly exceeds what traditional

[63] *Glimmerveen and Hagenbeek v The Netherlands* (1979) Application no. 8348/78.
[64] *Garaudy v France* (2003) Application no. 65831/01.
[65] *Chauvy v France* (2004) Application no. 64915/01 at para 69.
[66] *Ozgur Gundem v Turkey* (2000) Application no. 23144/93 at para 43; The Office of the United Nations High Commissioner for Human Rights, General Comment No. 34 (2011) at para 7, available at *http://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf*.
[67] European Parliament and Council Directive 2000/31/EC OJ 2000 L178, Art 15; Case 360/10 *Belgische Vereniging van Auteurs, Componisten en Uitgevers CVBA (SABAM) v Netlog NV* [2012] ECLI 85.
[68] *Salov v Ukraine* (2005) Application no. 65518/01.

legal systems are designed to withstand.[69] Additionally, the global nature of the internet is at odds with the "strictly territorial"[70] enforcement power of states, so where 'fake news' is published outside of the jurisdiction, national 'fake news' laws are powerless. Therefore, the law must rely on code, the hardware and software of the internet, to effectively regulate. The relationship between law and code will be explored further below.

(2) Media regulation

As the practices of traditional media are contributors to 'fake news', there requires special consideration of how the law can address 'fake news' as it relates to the media. In the UK, the press currently self-regulates its online and offline content, and its regulation includes the imposition of accuracy standards.[71] However, these accuracy standards only impose duties to take reasonable care to report accurately and do not amount to an obligation of truth. As such, where journalists have reported inaccurately, it is the measures they took rather than the accuracy of the information that is adjudicated. In theory, obligations under codes of conduct should inhibit journalists from crafting pieces so misleading that they misrepresent the truth. However, the absence of requirements to offer balanced news reporting and duties to guarantee pluralism as well as an online environment that prioritises clicks over content has resulted in the teetering of the characteristic partisanship of the British press into 'fake news' territory. Analysis of the one hundred most popular articles shared on social media about Brexit revealed that "the most popular dubious stories on British politics were almost always the work of long-established news outlets and relied at most on exaggeration rather than fakery".[72] This suggests that the press' duty to take care in reporting accurately is not being sufficiently enforced by the self-regulatory bodies. Therefore, one way in which the law could directly address 'fake news' is to inject itself into the regulation of the press through co-regulation. This is supported by the obligations of journalists under the ECHR. Whether this is desirable, especially in the UK where there is a long tradition of self-regulation, is outside the scope of this article, but it should be acknowledged that there are strong arguments in favour — notably, promoting the proper exercise of freedom of expression — and against — principally, retaining the independence of the press.

E. CAN LAW INDIRECTLY ADDRESS 'FAKE NEWS'?

This article has determined that the law can directly address 'fake news'. However, the characteristics of cyberspace make direct legal intervention of little effective use. This underlines the importance of considering indirect ways in which the law can act through the manipulation of other regulatory modalities. Code, in particular, can play the crucial role of an online enforcer. These modalities, and how the law can address 'fake news' through them, will be considered in turn. The scope of this discussion will focus on the online dimension of 'fake news' as, offline, the law remains an effective regulator.

(1) Code

Code dictates what one can and cannot do through physical or technical constraints. In the online context, code is the hardware and the software of the internet. As code has the ability to tackle the

[69] D Rowland, U Kohl and A Charlesworth, *Information Technology Law* (2017) 82; M Mueller, *Networks and States: The Global Politics of Internet Governance* (2010) 4.
[70] Rowland et al, *Information Technology Law* (n 69) 84.
[71] IPSO, Editors' Code of Practice (2017), available at *https://www.ipso.co.uk/editors-code-of-practice/*; IMPRESS, The IMPRESS Standards Code (2017), available at *https://impress.press/standards/impress-standards-code.html#show-2*.
[72] J Waterson, "Britain has no fake news industry because our partisan newspapers already do that job' (24 January 2017) BuzzFeed News, available at *https://www.buzzfeed.com/jimwaterson/fake-news-sites-cant-compete-with-britains-partisan-newspape?utm_term=.yn3a4XlaE#.pqRjw2QjG*.

scale of behaviour online and is not limited to geographic borders, it is the modality best suited to regulate cyberspace. Those that create code control code. Therefore, the creators of cyberspaces, like online platforms, and those who facilitate access to those spaces, the internet service providers (ISPs) are in the unique position to remove or block access to 'fake news' that law enforcers generally are not in. However, through encouraging controllers of code to aid in the battle against 'fake news' by the threat of sanction, the law can indirectly address 'fake news' through code. The ways in which the law can compel online actors will be considered individually.

(a) Limited liability laws

Intermediary liability laws[73] provide a mechanism to encourage online platforms' engagement by the threat of liability. Commonly known as 'Notice and Takedown' (NTD), intermediaries have to "expeditiously"[74] remove or disable access to content that they host after gaining actual knowledge to escape liability.[75] NTD duties only apply with regard to "unlawful activity and information";[76] therefore, intermediary liability laws can only be used as a tool of enforcement where there is prior "fake news" legislation in place. Although effective, NTD bears a high democratic cost.[77] In compelling online platforms to enforce the law, NTD creates an environment where public spaces are increasingly actively regulated by private actors. Whereas government enforcement is guided by democratic principles, balanced between the interests of the public and the interests of the state, and held accountable through checks and balances; private enforcement is guided by the principles of business, notably the maximisation of profit and the minimisation of liability. Indeed, NTD encourages overly-cautious[78] behaviour that, considering the many different types of 'fake news', could have a chilling effect on speech. However, the availability of algorithms[79] that can detect obvious markers of "fake news", while remaining sensitive to legitimate forms of speech, like satire and politically biased content, inspires confidence that online regulation will not be overly broad. Moreover, without public accountability, determinations of what constitutes 'fake news' are susceptible to being coloured by the biases of the controllers. Such concerns are not mollified by the use of algorithms as they are programmed by humans and are not value-free. Therefore, the democratic costs of NTD — the threat that it poses to the online exercise of freedom of expression — outstrips any justifications of efficacy or ease.

(b) Site blocking provisions

"Fake news" legislation can confer power onto courts to issue site blocking injunctions on ISPs. Through Cleanfeed, a content blocking system shared by major British ISPs, access to sites that are overwhelmingly dominated by 'fake news' could be effectively blocked. Section 97A of the Copyright, Design and Patents Act 1998 gives courts such power in relation to copyright infringement, and copyright holders have successfully relied on it.[80] A similar provision with respect to 'fake news' would be a legal tool to ensure code enforcement of the law. However, unlike

[73] The Electronic Commerce (EC Directive) Regulations 2002, SI 2002/2013; European Parliament and Council Directive 2000/31/EC (n 67).
[74] The Electronic Commerce Regulations 2002, Reg 19; European Parliament and Council Directive 2000/31/EC (n 67), Art 14.
[75] *Ibid*.
[76] The Electronic Commerce Regulations 2002, Reg 19.
[77] L Edwards, "The role and responsibility of internet intermediaries in the field of copyright and related rights" (2010) World Intellectual Property Organization at 11-12, available at *http://www.wipo.int/export/sites/www/copyright/en/doc/role_and_responsibility_of_the_internet_intermediaries_final.pdf*.
[78] *Ibid*.
[79] J Snow, "Can AI win the war against fake news?" (13 December 2017) MIT Technology Review, available at *https://www.technologyreview.com/s/609717/can-ai-win-the-war-against-fake-news/*.
[80] Copyright, Designs and Patents Act 1988, s 97A; *Twentieth Century Fox Film Corp & Ors v British Telecommunications Plc* [2011] EWHC 1981*; Dramatico Entertainment Ltd & Ors v British Sky Broadcasting Ltd & Ors* [2012] EWHC 268.

copyright, 'fake news' does not directly engage with rights or interests of others; as such, there is no natural base of claimants to rely on a site blocking provision. Therefore, the law would have to establish a government body tasked to address 'fake news' that can bring about actions. Additionally, site blocking injunctions are democratically attractive because they would require adjudication of whether content is 'fake news'. Caution should be had as they focus on the source of rather than individual pieces of content, so site blocking runs the risk of silencing sites that contain a mixture of 'fake news' and legitimate news. An approach similar to that of copyright law would cure such reservation with blocking injunctions only being issued where there is an overwhelming proportion of 'fake news' on the sites concerned. This would mitigate the adverse democratic effects of site blocking. Nonetheless, even though site blocking is democratically attractive in comparison to NTD, blocking injunctions are of questionable effectiveness, which ultimately renders it a poor enforcer of law. Adept users can simply use circumvention measures to access the site or turn to other replacement 'fake news' sites.[81] Therefore, intervention through site blocking would be fruitless.

(c) Direct regulation

Rather than using legal mechanisms to compel digital intermediaries, the law could directly regulate them. In particular, the law should regulate the controllers that are the closest to 'fake news': online platforms. Through direct regulation, the law could not oblige online platforms to generally enforce 'fake news' laws because Article 15 of the E-commerce Directive precludes general monitoring obligations.[82] However, although direct regulation cannot oblige general enforcement, the law can nonetheless target the environment that propagates 'fake news'. Specifically, the law could impose interoperability requirements on social networks that would allow users to easily switch between platforms. As it currently stands, the inability of users to easily transfer their data between profiles contributes to a 'lock in'[83] effect that makes online platforms immune to consumer disquiet. Interoperability requirements would enable consumer choice and create an environment where concerned users could choose platforms with stronger controls on 'fake news'. As a result of this choice, market forces could encourage platforms to take proactive measures against 'fake news'. Additionally, the imposition of media plurality and diversity standards could force online platforms to veer away from their practices of hyper-personalisation that create filter bubbles encouraging the presentation of 'fake news'.[84] Moreover, direct regulation of platforms could ensure that democratic standards of enforcement are adopted. Requirements for transparency and due process would alleviate the inherent democratic deficiencies of code enforcement.

(2) Market

The market is another regulatory modality through which the law can address 'fake news'. Advertising is a driver of the digital economy and is a central component of business models for online platforms that provide free services. The larger the audience and the more clicks, the more advertising revenue a site can generate.[85] This aspect of the digital economy has created a financial incentive for the creation of polarising, if not false, political content. Content that affirms political

---

[81] B Danaher, M D Smith and R Telang, "The effect of piracy website blocking on consumer behavior" (2015) Carnegie Mellon University's Initiative for Digital Entertainment Analytics (IDEA), available at https://www.thepriceofpiracy.org.au/content/The%20Effect%20of%20Piracy%20Website%20Blocking%20on%20Consumer%20Behvaiour.pdf.

[82] European Parliament and Council Directive 2000/31/EC, Art 15; *SABAM v Netlog* (n 67).

[83] M Eurich and M Burtscher, "The business-to-consumer lock-in effect' (2014) *Working Paper, University of Cambridge, Cambridge Service Alliance*, available at *https://cambridgeservicealliance.eng.cam.ac.uk/resources/Downloads/Monthly%20Papers/2014AugustPaperBusinesstoConsumerLockinEffect.pdf*.

[84] D Spohr, "Fake news and ideological polarization: filter bubbles and selective exposure on social media" (2017) 34(3) Business Information Review 150.

[85] UK Parliament, "Written Evidence Submitted by the Internet Advertising Bureau UK" (2017) at para 6, available at *http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/culture-media-and-sport-committee/fake-news/written/48224.pdf*.

views garners clicks; content that enrages garners clicks. Disruption of this enabling market could minimise these financial incentives and reduce 'fake news'. Therefore, the law can indirectly address 'fake news' by manipulating its market conditions. For example, a governmental body created by 'fake news' legislation could be tasked to identify and compile a blacklist of 'fake news' sites to issue to advertisers. The City of London Police's Intellectual Property Crime Unit has used this method to combat copyright infringing sites and has seen a "64% decrease in advertising from the UK's top ad spending companies"[86] on the relevant websites. Although promising, this drop does not represent decreases in revenue as smaller advertising companies can fill the vacuum. Nevertheless, blacklists for 'fake news' sites could prove effective. However, there are democratic reservations from using such blacklists; because of the political nature of the content of 'fake news', advertising blacklists could be seen as an indirect way of promoting government-approved media. Additionally, the law could impose transparency requirements on the advertising industry. As it stands, common industry practice, specifically programmatic advertising which uses algorithms for automatic targeted and personalised advertising, is notoriously opaque.[87] Transparency requirements would enable the growing body of advertisers that do not want to be associated with,[88] or fund 'fake news' to choose not to. This consumer choice could negatively affect the market for 'fake news' and be a financial disincentive.

(3) Social norms

Lastly, the law could combat 'fake news' through requiring promotion of online literacy in schools. A cyber-literate populace that can distinguish fact from falsity would not be as susceptible to believing 'fake news'. A shift of social norms could come through the introduction of laws that impose obligations for schools to provide online literacy classes. However, there would need to be a body created to design such classes so that teaching is standardised. Even though online literacy, in itself, is apolitical, there remains a risk that such classes could promote political biases. Therefore, the body responsible for the content of the classes must be independent from party politics, and the content should focus on basic analytic skills that equip students with the tools to make evaluative judgements of the veracity of the information online. A public with these skills would mitigate the potential harms of 'fake news'. This approach was adopted by the Italian government in 2017[89] and is an attractive long-term solution to 'fake news'. However, it should be noted that online literacy taught in formal education settings excludes large swathes of the populace that have exited school or university. As such, online literacy education initiatives would not be wholly effective. Nonetheless, it can shift the social norm surrounding online behaviour for current younger generations and those to come.

F. CONCLUSION

This article has found that there are many ways in which the law can address 'fake news'. The law can directly combat 'fake news' through clear and targeted legislation and media regulation. Although direct legal intervention does amount to a limitation of freedom of expression, the consequentialist tradition of the freedom provides a basis for the exception of 'fake news'. However, though the law

[86] City of London Police, "Operation Creative Sees 64 Per Cent Drop in UK Advertising" (2 March 2017), available at *https://www.cityoflondon.police.uk/advice-and-support/fraud-and-economic-crime/pipcu/pipcu-news/Pages/Operation-Creative-sees-64-per-cent-drop-in-UK-advertising-.aspx*.

[87] UK Parliament, "Written Evidence Submitted by News Media Association' (2017) at para 4, available at *http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/culture-media-and-sport-committee/fake-news/written/48244.pdf*.

[88] D Tambini (n 16); G Spanier, "Pressure on for clean-up in advertising's digital wild west" (21 February 2017) Evening Standard, available at *http://www.newsmediauk.org/write/MediaUploads/Fake%20News/Evening_Standard_22.2.17_Gideon_Spanier.pdf.*

[89] J Horowitz, "In Italian schools, reading, writing and recognizing fake news" (18 October 2017) The New York Times, available at *https://www.nytimes.com/2017/10/18/world/europe/italy-fake-news.html*.

can directly act, it cannot effectively act due to the nature of cyberspace. Additionally, the law can address 'fake news' indirectly through code, market, or social norms. Although regulation through code can cure the law's enforcement deficiencies, the involvement of private actors in the policing of public spaces raises significant concerns as to the democratic desirability of code regulation. Through regulation of the market, the law can undermine the financial incentives behind creating and disseminating 'fake news'. Lastly, the law can address 'fake news' by instigating a shift in social norms through the promotion of online literacy in schools.

With specific regard to the law addressing 'fake news' directly and indirectly through code, should the law even intervene? This article has not paid particular focus to such considerations, but a brief discussion is necessary to place the abilities of the law into context. Although there is justification and basis for exception within freedom of expression, 'fake news' is arguably not a matter of law. The law rarely intervenes with behaviour that fails to cause harm, either to another person or to oneself: criminal law is largely underpinned by the 'harm principle',[90] and tort law is characterised by harm done to others materialised in the form of damage.[91] Ultimately, although 'fake news' is harmful in undermining individual's autonomy and the foundations of democracy, it fails to cause harm in a tangible - and therefore actionable - way. Though discussion of which area of law is most suitable to address 'fake news' is outside the scope of this article, a broad consideration of these areas is nonetheless informative of the appropriateness of legal intervention. Moreover, one must consider the broader democratic implications of legal intervention against 'fake news'. Which is the greater risk to democracy: 'fake news' or measures to quell it? Both governments and populaces alike should be wary of introducing mechanisms that are arguably as threatening to democracy as 'fake news'. 'Fake news' legislation has the power to suppress freedom of expression as well as the power to promote it. Enlisting private actors to enforce laws without adequate democratic safeguards carries a similar risk of speech suppression. Ultimately, how 'fake news' is addressed should have as much of a prospective view as an immediate one. Who could be in power in the future and what they believe in is as relevant, if not more, than the political realities of today.

In conclusion, this article has illustrated that the law can in many ways address 'fake news' but that there needs to be particular consideration of the effectiveness of these methods and the associated democratic costs.

[90] J S Mill, "Of Individuality, as one of the Elements of Wellbeing" reproduced in J Grey and G W Smith (eds), *J S Mill's "On Liberty" in Focus* (2012) 72.
[91] J Law and E A Martin, *A Dictionary of Law*, 7th edn (2014).

WHO IS MY PARENT? IS THIS MY CHILD? A CRITICAL EXAMINATION OF THE BASE FOR ATTRIBUTING PARENTAL STATUS IN RELATION TO SURROGACY IN THE UK

Ciara McCartney[*]

A. INTRODUCTION

There are arguably few aspects of family law that are more important than having clarity when it comes to the parent-child relationship.[1] This view was recently shared by Sir James Munby in the English High Court, in light of a series of cases, which highlighted that the current rules allow children to be born into legal limbo, ultimately creating uncertainty as to who will be the child's parent. The prime example of this is the law surrounding surrogacy. The current law requires a transfer of parental status from the gestational mother to the commissioning party, known as the "intending parents", meaning that the child's life will commence with some form of litigation.[2] This arrangement is wholly dependent on the gestational mother granting consent to the transfer of status, given that surrogacy arrangements are not currently enforceable in the UK.[3] In addition to the inevitable state interference, the law arguably merits criticism for attributing status to the gestational mother in the first place, as she does not enter into the surrogacy arrangement with the intention of parenting the child, unlike the intending parents. This has been a hot topic for academics both across the UK and internationally, whose views will be visited throughout this article.

It is contended that a review of the current law and debate surrounding surrogacy arrangements in the UK provides a compelling argument for regulating the law in accordance with intentions. This is shown through an analysis of the current law and highlighting its tendency to disregard the intending parents, before demonstrating how this law is clearly detrimental to the operation of surrogacy as a whole. It can then be effectively argued that the solution to this is by automatically attributing parental status to the intending parents, rather than the gestational mother. The aim of this analysis is to find a solution which allows parties to utilise surrogacy in a way which allows them full confidence and certainty of the final outcome. This analysis weeds out contrasting criticism surrounding surrogacy, which has a tendency to focus on the transfer of status and the negative connotations this may have. These criticisms will not be ignored, but instead weighed against

[*] DipLP Student at the University of Edinburgh.
[1] *In the matter of the HFEA 2008 (Cases A, B, C, D, E, F, G and H)* [2015] EWHC (Fam) 2602 at para 3 per Sir James Munby. The title of this article is taken from a quote in this judgment.
[2] Human Fertilisation and Embryology Act 2008 s 33(1).
[3] Surrogacy Arrangements Act 1985 s 1A.

a collection of more compelling claims in favour of the intending parents. The article will then touch upon how any new law could operate in practice, with legislation being concluded as the best option. A more detailed overview of surrogacy in the UK is nevertheless required before any of the aforementioned arguments can be explored.

B. CURRENT LAW

Surrogacy is described as an arrangement in which a woman agrees to carry a child, with the view of handing him or her over to the intending parents upon birth.[4] Parental status is attributed to the surrogate, as the Human Fertilisation and Embryology Act 2008 (hereafter HFEA 2008) permits that the gestational mother, and nobody else, is to be treated as the child's mother.[5] Any partner of the surrogate will be the child's second parent,[6] meaning said partner will be required to consent to the final stage of the process along with the surrogate.[7] This consent will allow the transfer of parental status and legal rights from the surrogate mother (and any relevant partner) to the intending parents.[8] However, whilst they are not prohibited, these arrangements are not currently enforceable either.[9] This means surrogacy as a whole is an uncertain venture; it is wholly dependent on the surrogate's decision to grant consent for a parental order.

The HFEA 2008 offers a wider range of persons permitted to enter into surrogacy agreements; prior to this there was the "need" for a father,[10] ultimately meaning that surrogacy was only an option for heterosexual couples, who were also required to be married.[11] The removal of the aforementioned provisions requiring both a father,[12] and marriage,[13] means civil partners and cohabitants can now utilise surrogacy too.[14] However, whilst the HFEA 2008 achieved equality in terms of sexual orientation, it nonetheless missed an opportunity to revise and resolve the problem at the heart of surrogacy: the attribution of parental status and the preference for gestation over intention.

C. ISSUES ARISING FROM THE CURRENT LAW

The current law ignores the primary function of surrogacy arrangements: allowing the intending parties to parent the child born from the arrangement. The HFEA 2008 provides examples of other forms of artificial reproductive technologies (ART), namely artificial insemination by donor (AID) and in vitro fertilisation (IVF),[15] that are regulated in order to give legal effect to the parties' intentions. In such instances, parental status is attributed to the gestational mother who has sought AID/IVF treatment in order to fulfil her desire to become a parent;[16] the donor, who intends to donate rather than parent any child, is not attributed with any legal status,[17] and the mother can consent to another party being the child's legal father or second parent, provided he or she intends to be so.[18] The

[4] 1985 Act s 1.
[5] 2008 Act s 33(1).
[6] *Ibid*. s 35(1), s 42(1).
[7] *Ibid*. s 54(6)(a), (b).
[8] *Ibid*. s 54(1); though note that the intending father may be treated as the child's father at birth in the event the surrogate has given prior consent to this under s 36(1).
[9] 1985 Act s 1A.
[10] Human Fertilisation and Embryology Act 1990 s 13(5).
[11] *Ibid*. s 30(1)(a), (b), (c).
[12] 2008 Act s 14(2)(b).
[13] *Ibid*. s 36.
[14] *Ibid*. s 54(2)(a), (b), (c).
[15] See A Griffiths, J Fotheringham and F McCarthy, *Family Law*, 4th edn (2015) paras 3-60-3-63.
[16] 2008 Act s 33(1).
[17] *Ibid*. s 41.
[18] *Ibid*. ss 35, 36, 42, 43, 44.

HFEA 2008 was therefore enacted in a way that allows AID/IVF to function properly As sources have rightly pointed out, if intention was ignored, it would see thousands of estranged donors become legal parents, thus contradicting the nature of donation.[19] This addresses the biological aspect of parenthood and dismisses it as the determining factor when attributing parental status.[20] To say otherwise would be a step backwards, given that adoption has been a legitimate means of obtaining parental status for close to a century in the UK. On this basis, the present article will not distinguish between partial surrogacy, namely when a surrogate's egg and donor sperm is used, and full surrogacy, whereby the embryo is created using the gametes of the intending parents, or with either sourced from donors.[21] Following on from this, if the basis of attributing parenthood is not a matter of genetics, it begs the question as to what exactly is. Surrogacy is not regulated in a way that gives effect to intention in the same way as AID or IVF, which means it is currently giving effect to something else.

It is argued that IVF and AID remain regulated in accordance with intention because these modes of ART do not infringe public policy to the same extent as surrogacy; they are in line with the alleged "traditional" concept of families, in which the intending mother gives birth to the child.[22] The key difference between surrogacy and other ART is of course gestation. Attempts to justify why the law relating to surrogacy is therefore treated differently to AID and IVF can be found in both case law and literature.

The concept of pre-natal attachments formed between the surrogate and the child carries a lot of weight in the courts. *Re Baby M*[23] involved a surrogacy arrangement whereby the surrogate changed her mind, stating "it wasn't until the day I delivered her that I finally understood that […] I was giving away *my* baby".[24] The court found in favour of the surrogate on the basis it was in the "best interests" of the child to stay with her gestational mother; a reason that is not uncommon in the courts, and commonly justified by the idea of pre-natal attachment between the gestational mother and the child.[25] This justification is not persuasive in terms of giving effect to the child's best interests as it relies heavily on an emotional bond that can only be expressed by the gestational mother, rather than the child. Additionally, it is possible to argue that the intending parents have become equally, if not more, attached to the child. They brought about the child's existence. Although this notion would be no more convincing, should the court use it as a basis for giving effect to the best interests of the child, it does suggest that the reasoning behind current case law is biased towards the gestational mother.

This claim is supported by the case of *Re Evelyn*.[26] In this case, the child arising from a surrogacy arrangement spent the first eight months of her life with the intending parents, before the surrogate changed her mind and the court ordered the child to be returned to her. This case provides an example of how the argument relating to attachment and bonding might have worked in favour of the intending parents. However, the judge ruled in favour of the surrogate and took the view that the gestational mother would be better at helping the child deal with "psychological problems relating to abandonment and identity".[27] Although these problems relating to abandonment and identity may

[19] K Horsey, "Challenging presumptions: legal parenthood and surrogacy arrangements" (2010) 4(22) Child & Fam LQ 449 at 465-466, 2372; L B Andrews, "Beyond doctrinal boundaries: a legal framework for surrogate motherhood" (1995) 81(8) Virginia LR 2342 at 2362; R Probert, "Families, assisted reproduction and the law" (2004) 16(3) Child & Fam LQ 273 at 274.
[20] T Callus, "A new parenthood paradigm for twenty-first century family law in England and Wales?" (2012) 32 LS 347 at 356.
[21] See Griffiths et al, *Family Law* (n 15) para 3-65; see also Horsey (n 19) at 451.
[22] Probert (n 19) at 274; J Dewar, "The normal chaos of family law" (1998) 61(4) Mod LR 467.
[23] *Re Baby M* 109 NJ 396 (1988).
[24] See R Cook, "Donating Parenthood: Perspectives on Parenthood from Surrogacy and Gamete Donation" in A Bainham, S D Sclater and M Richards (eds), *What is a Parent? A Socio-Legal Analysis* (1999) 133.
[25] *Re TT* [2011] EWHC 33 (Fam); *Re Marriage of Moschetta 30 Cal Rptr 2d 893* (Cal Ct App1994); *Re P* [1987] 2 FLR 421.
[26] *Re Evelyn* (1998) 23 Fam LR 53.
[27] *Ibid*. at 67.

never arise, they were given greater weight than the inevitable bond that would have existed between the child and intending parents. This shows that case law is not consistent in providing protection for emotional attachments that form during surrogacy. Rather, the courts will often ignore these attachments when they favour the intending parents and instead base decisions on hypothetical presumptions, thereby upholding the view that the child is best placed with the gestational mother. Relating this back to the best interests of the child argument, and based on the aforementioned inconsistencies, it could therefore be argued that the best interests argument is a disguise used to ensure that parenthood is rooted with the gestational mother.

Following this, the courts have further awarded parental status to the gestational mother on the basis that forcing a surrogate to give up "her" child could cause her great psychological harm[28] and enforcing such arrangements results in the exploitation of the surrogate who is "vulnerable".[29] This basis overlooks the point that surrogacy arrangements will not necessarily involve a surrogate who is vulnerable, and also fails to consider how greater regulation of the institution would only help to eliminate the possibility of any exploitation. The exploitation argument primarily stems from payments given in surrogacy arrangements,[30] with academics stating that surrogates may be more likely to be manipulated and fail to assert their own interests adequately during the process.[31] However, this claim is ultimately weak in the UK due to the prohibition on payments beyond reasonable expenses,[32] and the author does neither seek nor recommend any change in relation to this.

The issue of expenses will be revisited, but what is clear is that the law relating to surrogacy is largely focussed on the gestational mother at the expense of the intending parents, who are the instigators of the process in the first place. If this instigation, followed by intention, is sufficient to obtaining parental status in cases of IVF and AID, why is this not the case for surrogacy?

Academics have noted that family law is designed to create an "automatic devotion" to our offspring, meaning we cannot comprehend ever abandoning them.[33] However, if society acknowledges and accepts that people are permitted *not* to parent in certain circumstances - giving a child up for adoption being the clearest contemporary example[34] - this attitude could extend to surrogacy and gestational mothers in the future. The opposing argument here is that it is wrong to allow surrogates to make voluntary decisions regarding parental status before the birth of the child as they may later regret this.[35] Few academics have considered a law that would automatically attribute parental status to the intending parents rather than require a transfer of status, which is the case at present.[36] However, it is arguably the case that such automatic attribution would work to eliminate at least some of the disapproval a surrogate may receive for giving up "her" child, since the child will never be regarded as hers to give up. In line with this, the surrogate's lack of expectations of raising the child means she will not bond with the child as if it was her own.[37]

---

[28] *Re Evelyn* (1998) 145 FLR 90; *Re Marriage of Moschetta* (1994) 30 Cal Rptr 2d 893; see also Horsey (n 19) at 462.

[29] Horsey (n 19) at 451; C E Schneider, "Surrogate motherhood from the perspective of family law" (1990) 13(1) Harvard Journal of Law & Public Policy 125 at 127-128; V E Munro, "Surrogacy and the construction of the maternal-foetal relationship: the feminist dilemma explained" (2001) 7 *Res Publica* 13 at 20-21.

[30] E Blyth, "'I wanted to be interesting. I wanted to be able to say I've done something with my life': Interviews with Surrogate Mothers in Britain" (1994) 12 Journal of Reproductive and Infant Psychology 189 at 190.

[31] Schneider (n 29) at 127.

[32] Human Fertilisation and Embryology Act 2008 s 54(8); Surrogacy Arrangements Act 1985 s 2(1).

[33] Schneider (n 29) at 130; Cook (n 24) at 135.

[34] G Douglas, "Assisted reproduction and the welfare of the child" (1993) 46(2) Current Legal Problems 53.

[35] M M Shultz, "Reproductive technology and intent-based parenthood: an opportunity for gender neutrality" (1990) Wisconsin LR 297 at 351.

[36] A E Stumpf, "Redefining mother: a legal matrix for new reproductive technologies" (1986) 96(1) Yale Law J 187 at 204; K Horsey and S Sheldon, "Still hazy after all these years: the law regulating surrogacy" (2012) 20(1) Medical Law Review 67 at 86.

[37] J Levitt "Biology, technology and genealogy: a proposed uniform surrogacy legislation" (1991) 25(3) Columbia Journal of Law & Social Problems 451 at 462.

Legislators have not appeared tempted by statistics that reveal only 4-5% of surrogates change their minds and do not follow through with the arrangements.[38] By removing their right to enter agreements without total commitment, these figures would only fall. Expanding on this adds a new dimension to the discussion. Not only is it important to establish who a parent is and why, but also *when* they become one. It may be asked why we should be concerned with the current law surrounding surrogacy if it only causes problems in 4-5% of cases, but these cases are a direct consequence of the main issue: the attribution of parental status is ascribed to someone else before it can go to the intending parents.

D. PROPOSAL FOR REFORM: INTENTION OVER GESTATION

As previously noted by Sir James Munby, the law requires a measure of certainty when attributing parental status in order to avoid today's children being born into legal limbo.[39] The current law in the UK surrounding surrogacy is rigid by attributing parental status to the gestational mother upon the child's birth, but nevertheless fails to grant certainty as to who will raise the child as his or her own. Allowing intending parents to enter into surrogacy arrangements without any guarantee that they will obtain parental status is said to be a deliberate attempt to discourage these parties from entering into such agreements.[40] However, with the increasing rate of infertility and the corresponding decrease in number of children available for adoption,[41] the law ought to develop accordingly and provide parties with a regulated institution that secures their intention to become parents, and surrogacy allows them to do so in a way that mimics natural conception as closely as possible.[42]

When considering intention as a means of governing the attribution of parental status, a good starting point is Stumpf's analysis which breaks the procreative process into four stages: the child initiating stage, preparation, gestation and the child raising stage.[43] At present, there is an imbalance in the emphasis placed on the pre-gestational stages and those which follow. With regards to the child initiating stage, the only active participants are the intending parents. Stumpf argues that this is when the child's existence begins, and the intending parents therefore deserve credit as the conceivers of this child by gaining relevant rights and obligations.[44] This theory became reality in the US case of *Johnson v Calvert.*[45] The intending couple entered into an agreement with a surrogate who was inseminated with both of the intending parents' gametes, but when the parties fell out, the surrogate sought declaration of parentage. Whilst the lower courts ruled in favour of the intending parents purely because of the biological connection, the California Supreme Court instead based its decision in their favour on their intention to procreate a child whom they could raise themselves. It was held that "but for" their intention, the child would not exist.[46] Should we therefore attribute parental status in accordance with causation?

The dissenting opinion in *Johnson v Calvert* put forward the equally compelling argument that the surrogate's performance was just as indispensable to the child's birth as that of the intending parents.[47] This is, of course, true, given that the surrogate is carrying out the function the intending

[38] The Brazier Report, Surrogacy: Review for Health Ministers of Current Arrangements for Payments and Regulation, Report of the Review Team (HMSO Cm 4068 1998) para 3.38; see also E Jackson, *Regulating Reproduction: Law, Technology and Autonomy* (2001) 281.
[39] [2015] EWHC (Fam) 2602 para 3.
[40] Douglas (n 34) at 637.
[41] P H Schuck, "Some reflections on the Baby M case" (1988) 76(5) Georgia Law Journal 1793 at 1796, 1802.
[42] *Ibid*. at 1797.
[43] Stumpf (n 36) at 193.
[44] *Ibid*. at 196.
[45] *Johnson v Calvert* (1993) 19 Cal Rptr 2d 494 5 Cal.
[46] *Ibid*. paras 16-17 per Panelli J.
[47] See Cook (n 24) at 135.

parents cannot perform themselves. However, the intending parents would not have entered into the agreement had the surrogate stated an intention to raise the child, which is ultimately why more weight was given to the intending parent's indispensable role in the case. We can therefore conclude that attribution of parental status here was based not only on the intending parents' intention, but also on the lack of intention of the surrogate. Does this mean that a shift in the intention of the surrogate should have no impact under the proposed method of attributing parenthood?

The aforementioned 4-5% of problem cases that arise from surrogacy arrangements will most likely involve a surrogate who has changed her mind and decided she would like to parent the child.[48] This has been cited by some as a reason for posing a blanket ban on surrogacy, on the basis that a surrogate will not correctly anticipate her feelings towards "her" child.[49]

Using intention as a means for attributing parental status would mean the surrogate would be required to make an informed decision before entering into the agreement, in line with Schuck's proposed regulations concerning surrogacy.[50] Of course, making an informed decision does not eliminate the risk that a surrogate may change her mind later on, but doing so does not render her earlier decision involuntary. As Schuck notes, "second thoughts following difficult choices are ubiquitous in life", but that this is the risk we take when making commitments.[51] It therefore follows that the solution to the problems surrounding surrogacy is not to implement a blanket ban or attribute status to the surrogate in case she changes her mind, but rather to impose automatic attribution of parental status to the intending parents. In doing so, it is anticipated that the number of problem cases will decrease, as women who are more likely to change their minds during gestation will not volunteer to become surrogates, and those who do volunteer will do so on the basis they are simply a substitute for part of the process.[52] Such a premise is already reflected in medical literature, with Sokoloff strictly referring to surrogates as "donors".[53] As a result of this we could even see an increase in the number of women willing to be surrogates, encouraged by the certainty that they will never be legally responsible for the child.

This premise relies upon any second thoughts of the intending parents having no impact on their legal status that flows from surrogacy. It may be less likely given that they initiated the proceedings in the first place, but intending parents may too change their minds. A recent case making the headlines involved an intending father who "abandoned" his child born from a surrogacy arrangement upon discovering that "baby Gammy" had Down syndrome.[54] It would not make sense to reform the law in such a way that afforded parental status to the intending parents to reflect their initial desire, only to then force the surrogate to make decisions regarding the child's welfare in the instance that the intending parents change their minds. Forcing either party to take the child would hold them to higher standards than persons in instances of natural conception, who are free to place children up for adoption.[55] But the automatic attribution of parental status to the intending parents, rather than any transfer of such status, would mean that they were responsible for decisions regarding the child. The law should therefore be regulated accordingly. As Stumpf states: "the initiating parents

[48] However, the case of *Re G (Surrogacy: Foreign Domicile)* 2007 EWHC 2814 highlights the difficulty that may arise due to the Human Fertilisation and Embryology Act 2008 s 35(1) and s 42(1).
[49] Schneider (n 29) at 127.
[50] Schuck (n 41) at 1805.
[51] *Ibid*. at 1799.
[52] Horsey (n 19) at 469.
[53] B Z Sokoloff, "Alternative modes of reproduction: the effects on the child" (1987) 26(1) Clinical Pediatrics 11 at 13.
[54] K Marks, "Baby Gammy: Australian father who abandoned Down syndrome surrogate child now tries to access funds donated for his care" (19 May 2015) The Independent, available at http://www.independent.co.uk/news/world/australasia/baby-gammy-australian-father-who-abandoned-down-syndrome-surrogate-child-now-tries-to-access-funds-10261916.html.
[55] Stumpf (n 36) at 204.

should be responsible for a child that no one wants, and they should be entitled to a child that everyone wants".[56]

Moreover, it is unsurprising that academics are sometimes tempted to compare surrogacy to adoption. Those who do so often base proposals on the fact that parents are not allowed to give up parental rights until the child is born, allowing the gestational mother time to debate her much anticipated consent.[57] It is clear that a law structured around intention would not give the gestational mother any parental status to deliberate, as automatic attribution to the intending parents would be part-and-parcel of this, making the comparison with adoption irrelevant for current purposes. In order to avoid any transfer of status, and thus the stigma that flows from it, attribution of parental status to the intending parents would have to be upon conception rather than upon birth, as a child cannot be "parentless" throughout gestation.[58] This would therefore weaken the overall arguments in favour of gestational motherhood and strengthen the base for attributing parental status both automatically and in accordance with intention. Automatic attribution of parental status to the intending parents will also avoid the child's life beginning with excessive state intervention,[59] reflecting the importance of being able to determine parenthood with ease.[60]

This automatic attribution of status would also have benefits elsewhere, such as the need to apply for a parental order. Parental orders have been criticised for being "overly restrictive" and "burdensome and complex" on the basis that, for example, they are granted by way of litigation, and only where there is a biological connection to one of the intending parents, and the child is already living with the intending parents.[61] Removing parental orders from surrogacy will, in turn, remove the barriers associated with them.

E. ENCOMPASSING A WIDER GROUP OF PEOPLE?

Despite the HFEA 2008 addressing equality with regards to sexual orientation, it nevertheless requires at least one of the intending parents to be biologically related to the child.[62] This means that surrogacy fails to be an option when both of the intending parents are infertile.[63] Schuck argues that this should remain in place, on the basis that having one biologically related parent would decrease the odds of the child being abandoned,[64] although the aforementioned case involving Baby Gammy suggests otherwise.[65] If relying on biology has already been dismissed as a determinative factor for attributing parental status,[66] it should not be a pre-requisite for accessing surrogacy. In line with *Johnson v Calvert*, attributing parental status in surrogacy should be pushed to see past any genetic link, or lack thereof, and instead look to the parties' intentions. This will treat claims to parenthood equally.[67]

However, transcending the law of surrogacy does not end by removing biological requirements. The HFEA 2008 provides that the law will not grant a parental order in favour of single

[56] *Ibid*. at 205.
[57] A M Capron and M J Radim, "Choosing family law over contract law as a paradigm for surrogate motherhood, in surrogate motherhood" (1988) 16(1-2) Law, Medicine & Health Care 34 at 35; Douglas (n 34) at 636.
[58] Horsey (n 19) at 458.
[59] J L Dolgin, "Status and contract in surrogate motherhood: an illumination of the surrogacy debate" (1990) 38(2) Buffalo LR 515 footnote 67.
[60] *Re A* [2015] EWHC (Fam) 2602 at 3.
[61] Horsey & Sheldon (n 36) at 81-82.
[62] Human Fertilisation and Embryology Act 2008 s 54(2) limits parental order applications to (a) husband and wife (b) civil partners of each other, or (c) cohabitants.
[63] Horsey & Sheldon (n 36) at 82.
[64] Schuck (n 41) at 1805.
[65] Marks, *Baby Gammy* (n 54).
[66] Callus (n 20) at 356.
[67] Douglas (n 34) at 638.

persons,[68] as highlighted in recent media coverage.[69] This contradicts the current law regulating AID and IVF, under which the removal of the need for a father means that a woman can be the sole parent of a child born as a result of either process.[70] Only allowing surrogacy for couples was allegedly based on the "magnitude" of surrogacy and the idea that couples are better equipped to deal with this,[71] but it is not clear how this conclusion was reached. Had the law on AID and IVF not contradicted this premise, one could have assumed that surrogacy was regulated in such a way to avoid children being born into lone-parent families. But this doesn't stand if other forms of ART can operate contrary to this.

Various academics support the claim that contrasting laws concerning single parents in AID or IVF and surrogacy cannot co-exist.[72] It has been noted that attributing parental status to a single woman who has given birth as a result of AID or IVF, whilst prohibiting the attribution of parental status to a single man who has commissioned a surrogate, suggests that the law is stating that it would rather have single mothers than single fathers.[73] McCandless and Sheldon rightly conclude that the solution to this problem is granting equality and equating the law of surrogacy to that concerning other forms of ART. After all, it makes sense that a single person, male or female, may desire to be a parent and thus look to the law to act on this intention. Celebrities like Ricky Martin and Perez Hilton highlight the US approach to surrogacy and its willingness to grant parental status to intending parents, regardless of their single status.[74] In line with US law, and in light of the Office for National Statistics revealing that close to two million dependent children in the UK were raised by lone-parents in 2015,[75] the law on surrogacy should reflect the normality of lone-parentage. Access to surrogacy for single parents, as well as those without biological ties, would therefore flow as a direct benefit from regulating the law according to intention.

This issue was recently debated in the High Court in England, in a case which involved a single man who had commissioned an American surrogate, and later went on to raise the child as a lone parent.[76] The single father was, and still is, unable to obtain a parental order in the UK. However, the court did hold that the law was incompatible with the Human Rights Act,[77] and the Department of Health has stated that the government will be looking to update the legislation concerning this.[78] Whilst it may be anticipated that this particular issue regarding parental orders and lone parents may be resolved before surrogacy as a whole is reformed, this discussion highlights another aspect of the current law that is not fit for purpose.

Moving forward, whilst the proposed law would encompass a wider range of persons, meaning children could be born into families with only one parent, it follows that the law must also be clear on the maximum number of parents a child may have. Horsey notes that the way in which surrogacy operates could suggest a child may have up to six potential "parents": two gamete

---

[68] 2008 Act s 54(2).
[69] J Manel, "Father's hopes for single parent surrogacy law change*"* (17 August 2016) BBC News, available at http://www.bbc.co.uk/news/uk-37108407.
[70] 2008 Act s 13(5).
[71] Horsey & Sheldon (n 36) at 83.
[72] *Ibid*. at 83; J McCandless and S Sheldon, "No father required? The welfare assessment in the Human Fertilisation and Embryology Act 2008" (2010) 18(3) Feminist Legal Studies 201 at 213.
[73] *Ibid*.
[74] D. Kambouris, "Perez Hilton welcomes baby girl via surrogate" (11 May 2015) CBS News, available at http://www.cbsnews.com/news/perez-hilton-welcomes-baby-girl-via-surrogate/; S Palomares, "Ricky Martin finally opens up about his surrogate mom" (26 September 2014) Mamas Latinas, available at http://quemas.mamaslatinas.com/parenting/130130/ricky_martin_finally_opens_up.
[75] Office for National Statistics, "Statistical Bulletin: Families and Households: 2015" available at https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/families/bulletins/familiesandhouseholds/2015-11-05#lone-parents.
[76] *In the matter of Z (A Child) (No 2)* [2016] EWHC 1191 (Fam)
[77] Human Rights Act 1998.
[78] Manel, *Father's hopes* (n 69).

providers, a gestational mother and her partner and two intending parents.[79] Given the policy debate circling multi-parent families,[80] and the pre-existing stigma associated with surrogacy, a law based on intention should only be implemented insofar as it attributes parental status to a maximum of two persons.[81] By doing so, it equates parents and children from surrogacy to recognised and accepted family forms. A refusal to recognise both the surrogate and the intending mother as parents of the child was displayed in the aforementioned case of *Johnson v Calvert.*[82] This was to give effect to the intentions of the parties, but also with respect to policy considerations concerning the child himself, a party whose feelings cannot be overlooked in the current discussion.

F. SURROGACY FROM A CHILD'S PERSPECTIVE

Looking back to the case of *Baby M*, it was noted that the courts suggest children born from surrogacy arrangements were better left in the care of the gestational mother.[83] Additionally, in the aforementioned case of *Re Evelyn,* it was similarly held that any short term effects caused by removing the child from the care of her intended parent were overcome by the advantages of being placed with her gestational mother.[84] Mitchell states that the surrogacy debate has overlooked a crucial solution in this area: to involve the child and consider his or her needs in the agreement.[85] The "best interests" threshold is regularly used as a means for determining outcomes in family law generally and the United Nations Convention on the Rights of the Child (UNCRC) provides that the welfare of the child ought to be a primary consideration.[86] The Children Act 1989 provides that the welfare of the child is to be the court's paramount consideration when determining any questions relating to his or her upbringing,[87] with the Children (Scotland) Act 1995 repeating this requirement in relation to parental responsibilities.[88] Lord MacDermott notes that this threshold operates in a manner that prevails over all other considerations,[89] showing that the best interests standard has clearly asserted its overall position in family law.

Sokoloff questions whether ART should operate to fulfil one person's happiness by helping to conceive a child that may not be destined for happiness itself.[90] Does it therefore follow that the best interests standard ought to govern the attribution of parental status in a surrogacy setting? Claims suggesting this, including Mitchell's, suggest that the best interests could be discerned through an independent counsel who are responsible for gathering evidence in relation to the child's interests.[91] However, what Mitchell views as the "best interests" of the child is unclear, as although he states that these should not conflict with the parents' interests, he does not state what they actually are. He does state that the independent counsel could ensure that the child's rights are not violated,[92] but this would be unhelpful in the event that neither the gestational mother nor the intending parents seek to violate the child's rights, which is nearly always the case. His argument is less compelling when one

[79] Horsey (n 19) at 451.
[80] See J Wallbank and C Dietz, "Lesbian mothers, fathers and other animals: is the political personal in multiple parent families?" (2013) 25(4) Child & Fam LQ 451.
[81] It is also a requirement that these persons not be within the prohibited degree of relationships applicable to family law; e.g. 2008 Act s 54(2)(c); See also McCandless & Sheldon (n 73) at 198.
[82] Douglas (n 34) at 637.
[83] *Re Baby M* 109 NJ 396 (1988).
[84] *Re Evelyn* (1998) 23 Fam LR 53.
[85] W Mitchell, "Surrogate parenthood – an analysis of the problems and a solution: representation for the child" (1986) WM Mitchell LR 143 at 173.
[86] United Nations Convention for the Rights of the Child art 3.
[87] Children Act 1989 s 1(1).
[88] Children (Scotland) Act 1995 s 1(1); See also The Commission of European Family Law and its Principles Regarding Parental Responsibility principle 3:3.
[89] *J v C* [1970] AC 668 at 710-711.
[90] Sokoloff (n 53) at 14.
[91] Mitchell (n 85) at 176.
[92] *Ibid*. at 176-177.

considers that it confuses the ethical framing of ART,[93] with Solberg stating that the overall goal of ART is to deliver a baby and make people parents in order for them to experience family life with children. Solberg further argues that this is not confined to biology, adding support to the above discussion.[94] Therefore, with the overall aim being to produce functioning families,[95] it follows that ART merits the attribution of parental status to any *social* parents: those combining "motivation, intention, involvement and nurturance".[96]

Does this mean the "best interests" threshold is irrelevant as far as surrogacy is concerned? Stumpf emphasises that the welfare of children is promoted by a system that attributes parental status as unambiguously as possible.[97] This notion discredits the best interests threshold, not only on the basis that it produces unpredictable results, but it also encourages judicial intervention at the early stages of a child's life and prevents any certainty in the law with regards to who a parent is. It is for this reason that academics have posed a different, but by no means inferior, standard to govern surrogacy and ART in order for it to operate effectively and act on intentions: the standard of "adequate care".[98] This is in accordance with the idea that by creating children through surrogacy, it will not necessarily result in overall harm to the child, in the same way that natural conception will not guarantee the child's best interests.

If surrogacy is not about choosing the "best" parents for a child, the proposed reforms to attribute parental status based on intention can operate with relative ease. Like natural parents, the state should intervene at any point when a child is at risk of significant harm, but not necessarily before parental status is even attributed. This intention-based model will not stray any further than current conventional rules, where children are not given a say on who their parents will be.[99] While such a model will therefore grant certainty as to who a child's parent is, it will not go as far as to guarantee that this child will not remain affected by any stigma towards surrogacy, including the idea of abandonment described by the judiciary in *Re Evelyn*.[100] However, this stigma can be reduced by replacing the current law requiring a surrogate to "give up" the child with the aforementioned automatic attribution to the intending parents,[101] as she would effectively be "womb leasing"[102] and handing over what was never hers.

G. CONTRACT VS. LEGISLATION

The certainty offered by this proposed reform must be able to translate into practice, necessitating a brief discussion as to how this attribution of parental status according to intention will operate, either by way of contract or legislation. Shultz explains that when intentions meet policy criteria, these intentions can become legally enforceable under the rules of contract law.[103] Douglas then takes this premise and injects it into a surrogacy setting, stating that a contract is the simplest way to prove intentions and thus attribute status accordingly.[104] However, this alleged simplicity doesn't necessarily render contract as the appropriate method. Contracts offer a relative amount of flexibility in terms of drafting, which may in turn produce blurred lines as to what will and will not infringe public policy. A

[93] B Solberg "Getting beyond the welfare of the child in assisted reproduction" (2009) 35(6) Journal of Medical Ethics 373.
[94] *Ibid*. at 374.
[95] *Ibid*. at 373-374.
[96] Cook (n 24) at 135.
[97] Stumpf (n 36) at 204.
[98] *Ibid*. at 207; Solberg (n 93) at 373.
[99] Shultz (n 35) at 341.
[100] See also Schneider (n 29) at 130.
[101] *Ibid*. at 127.
[102] Douglas (n 34) at 641.
[103] Shultz (n 35) at 346-347.
[104] Douglas (n 34) at 640.

contractual approach thus weakens the argument that regulating the law to give effect to intention will avoid a child's life beginning with litigation. Epstein argues in favour of full contractual enforcement, but rightly says that this should not stretch so far that it encompasses the right on the intending parents to order an abortion or the prohibit one,[105] on the basis that this choice is "so visceral and personal" that it cannot be subject to the ordinary rules of contract.[106] Allowing contracts to invade this area of family law runs the risk of parties attempting to enforce inappropriate terms. Andrews states that the answer therefore lies in legislation.[107] By going down the route of amending state parentage statutes, legislators are forced to address all of the aforementioned legal issues and allow the new law to operate unambiguously, similar to the law governing AID and IVF. When addressing issues in the current law however, legislators may be reluctant to abolish one particular aspect: namely, the prohibition of monetary payments beyond reasonable expenses.[108]

H. SURROGACY AND MONEY

Looking at surrogacy in a commercial light for the present purpose poses one question only: should we be able to pay for parental status? According to academics, the primary reason against permitting payments is that this reduces people to commodities;[109] it is said to be a policy measure in order to protect "poor, single women" from being exploited and to prevent "baby selling".[110] Reversing the law to allow persons to pay for parental status would therefore attract the same stigma that the proposed reform seeks to fight. This could be counter-argued by research stating that allowing women to exercise their autonomy and thus agree to be a surrogate upon stipulation of a fee should be regarded as empowering rather than exploitative,[111] with *Baby M* said to display a payment for the surrogates "service" rather than the "sale" of the baby.[112] However, weighing up the risks and benefits of both options leads to the conclusion that the current law regarding payments should prevail. The likely public condemnation for allowing surrogacy to operate on a commercial basis would cast a shadow on the otherwise positive reforms regarding the attribution of parental status. New legislation should therefore be given effect through a non-profit organisation, keeping criticism concerning exploitation at bay.[113] The Warnock Committee ironically rejected the idea of regulating a limited, non-profit service on the basis it would encourage people to use surrogacy,[114] showing legislators are also of the impression that this is an effective means of implementing the proposed reform.

I. CONCLUSION

In summary, the current law renders surrogacy an uncertain venture, wholly dependent on the co-operation of the gestational mother. The current attribution of parental status to the gestational mother gives rise to the surrounding stigma that comes with surrogacy. Both of these defects can be ameliorated by automatically attributing parental status to the intending parents, in line with the law on IVF and AID. Implementing this proposal means surrogacy can operate in a way that gives effect to the overall purpose of making the intending parties parents in order for them to experience family

[105] R Epstein, "Surrogacy: the case for full contractual enforcement" (1995) 81 Virginia Law Review 2305 at 2336.
[106] M Field, *Surrogate Motherhood: The Legal and Human Issues* (1990) 25.
[107] Andrews (n 19) at 2372.
[108] HFEA 2008 s 54(8); The Surrogacy Act 1985 s 2(1).
[109] Schneider (n 29) at 128; Dolgin (n 60) at 548.
[110] Capron & Radim (n 57) at 36.
[111] Horsey (n 19) at 451.
[112] Capron & Radim (n 57) at 36.
[113] Dolgin (n 59) at 549.
[114] The Warnock Report (Report of the Committee of Inquiry into Human Fertilisation and Embryology) 1984, para 8.18; See also Jackson (no 38) at 281.

life with children. It will do so without the same level of stigma attached to the law at present, as removing the current transfer of parental status will thus work towards diminishing the contradicting connotations of the child being "given up" by his mother, as well as this baby being ripped from the arms of a mother who has bonded with him or her during gestation. This provides the clarity that the current law is lacking. It will also encompass a wider group of people to reflect modern-day families, often absent of biological connections or more than one parent. This reform is best implemented through legislation, in order to prevent parties having too much flexibility with regards to the surrogacy arrangements, but also to force legislators to take ownership of the problems with the current law. The monetary aspect of surrogacy should, however, remain the same as the present law provides, in order to maintain the overall positive theme of the proposed law.

SUBMISSIONS

We are always keen to receive submissions from students, both undergraduate and postgraduate. Below are some basic guidelines for contributors as well as suggestions for submissions.

GUIDELINES

The Editors are pleased to consider:

Articles (up to 10000 words)
Shorter Analysis Pieces
Case Notes
Book Reviews
Reports

It is not necessary for articles to conform strictly to the specified word limit. However, authors should ensure that all submissions are as concise as possible and conform to the House Style, which is available from our website. While we encourage authors to write new material for the *Edinburgh Student Law Review*, we are also happy to accept pieces written in the course of studies.

For further information, please refer to our website: http://www.eslr.ed.ac.uk/.