

Manchester Review of Law, Crime and Ethics

The Student Journal of
The University of Manchester, School of Law

VOLUME VI
OCTOBER 2017

The Manchester Review of Law, Crime and Ethics is an annual, student-led, peer-reviewed journal founded at The University of Manchester, School of Law. The journal exhibits the best academic work in law, criminology and ethics, at the undergraduate and postgraduate levels.

VOLUME VI

OCTOBER 2017

Manchester Review of Law, Crime and Ethics

© *Manchester Review of Law, Crime and Ethics*, 2017

All rights reserved.

The *Manchester Review of Law, Crime and Ethics*, its editors, and The University of Manchester, School of Law are not responsible for the opinions expressed in this journal.

No part of this journal may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without the prior express written consent of the *Manchester Review of Law, Crime and Ethics*, The University of Manchester, School of Law or as expressly permitted by law.

All enquiries should be addressed to:

Manchester Review of Law, Crime and Ethics

The University of Manchester
Williamson Building, School of Law
Oxford Road
Manchester M13 9PL
United Kingdom

review.manchester@gmail.com

Typeset in Times New Roman
Printed in the United Kingdom

ISSN 2399-4630 (Print)
ISSN 2399-4649 (Online)

This year's volume was
generously funded by The
University of Manchester,
School of Law.

The Editorial Board is grateful
for the School of Law's
continued support.

Manchester Review of Law, Crime and Ethics

Editorial Board

Editor-in-Chief
Rohan Shah

Deputy Editor-in-Chief
Dhruv Bhatia

Editors
İşıl Aral
Tanjia Bashor
Claire Ceney
Gan Ming Hui
Marina Iskander
Isabella Kirwan
Fiona Long
Abisola Odusanya
Kevin Patel

Honorary Advisory Board

Lesley Anderson QC
HH Judge Nigel Bird
Dr. Nicholas Braslavsky QC
Dinah Crystal OBE
Paul Lowenstein QC
Gerard McDermott QC
Dr. Margaret Mountford
Toby Seddon

Preface from the Head of the School of Law

It is my pleasure to welcome the sixth edition of this review. What started out several years ago as an idea from the student body has now become a feature of the Law School which is developing an increasing reputation in the field. It covers the full range of areas of our work and involves students on all programmes.

The Law School is a community of scholars dedicated to the pursuit of higher learning about justice and rights in society, through research and education. The papers in this volume show the keen interest of our community in exploring cutting-edge issues and reflect the intellectual diversity and inter-disciplinarity that is a distinctive feature of the School. I hope all readers find something of interest and many in our community are encouraged to contribute in the future.

The authors are to be congratulated, as is the editorial team who have put in many hours of work. As Head of School, I am very proud that this review is continuing to thrive. I look forward to reading many more future volumes.

Professor Toby Seddon
September 2017

Preface from the Editor-in-Chief

I am delighted to present the culmination of many months' worth of hard work by our Editorial Board: Volume VI of the Manchester Review of Law, Crime and Ethics. What started in 2012 as a little-known, yet ambitious student publication, has quickly blossomed into a peer-reviewed and respected academic journal. This is attributable not only to the excellent work of students at the University of Manchester School of Law, but also the commitment of its many distinguished scholars.

Over the past year, we have sought to expand the Review's reach and strengthen its reputation. Firstly, we have continued to host free online versions of the Review on the School of Law's website and continued our relationship with HeinOnline. Secondly, we have obtained an ISSN and made the Review available at The British Library. Finally, we have engaged in an aggressive social media campaign to promote our scholarship and brand across Facebook and Twitter, forging links with other law schools, journals, and lawyers across the world.

Of course, none of this would be possible without the continued support of the School of Law. I would also like to thank the Editorial Board for their dedication, perseverance, and hard work. Finally, thank you – our readers – for continuing to engage with our work.

Lastly, I would like to introduce my successor, Kevin Patel, who I am certain will continue the Review's tradition of academic excellence.

Rohan Shah
August 2017

Table of Contents

Mass sexual violence in Darfur: An analysis of macro-, meso-, and micro-level variables.....	8
<i>Amy Cross</i>	
Opting out of the CISG: is the law too complicated for such a popular choice?.....	23
<i>Océane Girard</i>	
Class action (not US-style): enhancing access to justice.....	49
<i>Shinthean Ng</i>	
Dworkin's Hercules as a model for judges.....	58
<i>Arvinth Rai</i>	
Pregnant women may have moral obligations to foetuses they have chosen to carry to term, but the law should never intervene in a woman's choices during pregnancy.....	69
<i>Elizabeth Chloe Romanis</i>	
Should the law recognise rights in all humans and only in humans?.....	86
<i>Aneela Samrai</i>	
An exploration into the factors shaping victim reporting of partner abuse to the police.....	95
<i>Victoria Smith</i>	
The regulation of the civilian and military realms of cyberspace.....	121
<i>Jake Taylor</i>	
A critical assessment of the separability doctrine, its impact and application.....	148
<i>Fiona Winnifred Lakareber</i>	

Mass sexual violence in Darfur: an analysis of macro-, meso-, and micro-level variables

Amy Cross¹

Mass sexual violence is a common phenomenon during conflict, occurring across the world and historically over time. Many theories of why it occurs have been argued in the criminological literature, many focusing on group processes, and others on individual perpetrators. This paper analyses the incidence of mass sexual violence in Darfur, Sudan. It discusses the interplay of broad, macro-level variables with meso- and micro-level variables and finds that it is not one theory that explains mass sexual violence in this context, but an interaction and compounding of many different variables, all of which contribute to the occurrence of the phenomenon.

It is recognised within the literature that mass sexual violence has consistently been a feature of conflict both temporally and geographically,² and it has been a feature of the genocide in the Darfur region of Sudan since the onset of the conflict.³ While criminology has had relatively little engagement with sexual violence in conflict, there are some theories which explain sexual violence in conflict, and others which explain other forms of crime but can be applied to this type of offending. This essay will aim to distinguish some of these theories in terms of their level of analysis, to apply them to this context, and to analyse the interaction of macro, meso and micro factors that have influenced the problem. Macro levels of analysis will be discussed through the lens of Staub's equation of difficult life conditions and cultural preconditions resulting in mass violence,⁴ including the concepts of patriarchy and

¹ LL.M. Candidate, The University of Manchester, School of Law.

² Susan Brownmiller, *Against Our Will* (Penguin 1975) 31-32; Eileen Zurbriggen, 'Rape, War and the Socialisation of Masculinity: Why Our Refusal to Give Up War Ensures That Rape Cannot be Eradicated' (2010) 34 *Psychology of Women Quarterly* 538.

³ John Hagan and Wenona Rymond-Richmond, *Darfur and the Crime of Genocide* (Cambridge University Press 2009).

⁴ Ervin Staub, *Overcoming Evil: Genocide, Violent Conflict, and Terrorism* (Oxford University Press 2011).

hegemonic masculinities. Meso levels of analysis will focus primarily on ideology which is an important factor in joining the macro and micro together to create a multi-layered theoretical framework in which to explain sexual violence in conflict. Finally, this essay will analyse some of the micro level theoretical frameworks regarding individual and group explanations for the committal of such atrocities. This essay will posit that although all three levels individually are influential, it is the interplay of the three levels which creates conditions in which the problem of sexual violence can thrive. Firstly, though, as crime is a social construction, it is important to position sexual violence within conflict in a legal framework in order to identify it as a crime.

Although widespread sexual violence has been recognised in Darfur, it has been difficult to quantify due to a lack of reporting, however, the United Nations documented 117 incidents in 2014.⁵ It is estimated that this figure is underestimated, owing to underreporting due to fear of further victimisation from the state, and the stigma attached to sexual violence in this setting.⁶ The Convention on the Prevention and Punishment of the Crime of Genocide defines genocide as acts intended to cause the destruction of whole or part of a group.⁷ Although the definition does not specify sexual violence, it does include causing bodily or mental harm to members of the group or preventing births within the group, both of which can include sexual violence. Furthermore, the Akayesu case ruled that rape constitutes part of genocide and should be considered a war crime and a crime against humanity.⁸ Hagan and Richmond-Rymond suggest the extent of the problem in Darfur is sufficient enough to prosecute and this has been confirmed by the International Criminal Court (ICC) which has issued arrest warrants

⁵ UNSC Report of the Secretary General on Conflict-Related Sexual Violence (2015) UN Doc S/2015/203.

⁶ *ibid*; John Hagan, Wenona Rymond-Richmond and Alberto Palloni, 'Racial Targeting of Sexual Violence in Darfur' (2009) 99 American Journal of Public Health 1386.

⁷ Convention on the Prevention and Punishment of the Crime of Genocide (9 December 1948) 78 UNTS 277.

⁸ *Prosecutor v Akayesu* (Judgment) ICTR-96-4-T, T Ch I (2 September 1998).

for a number of individuals in Sudan, including President al-Bashir, for crimes including rape.⁹

Macro levels of analysis address the deep history and wider cultural and economic context of an incidence of mass violence. Staub theorised that mass violence follows a combination of factors, in particular, difficult life conditions and cultural preconditions which result in psychological and social reactions.¹⁰ Both of these factors were present prior to the conflict in Darfur. The region was in drought and suffered extreme poverty, there was competition for land and resources, and the government had neglected the region during and since Sudan had gained independence from British colonisation.¹¹ Difficult life conditions including structural inequality alone have been used by strain theorists to explain crime in peacetime,¹² and in countries where gendered structural inequality is particularly prevalent, conflict and sexual violence against women and girls is more common.¹³ However, in applying Staub's theory, difficult life conditions alone are not enough to explain mass sexual violence.¹⁴ There are countries and regions across the world whose populations suffer difficult life conditions but are not subject to mass sexual violence, and likewise, there have been incidences of mass sexual violence in conflict, such as the war in Bosnia, in which difficult life conditions were not particularly relevant. Therefore, cultural preconditions should also be considered in relation to the problem.

⁹ John Hagan and Wenona Rymond-Richmond, *Darfur and the Crime of Genocide* (Cambridge University Press 2009); *Prosecutor v Omar Hassan Ahmad Al Bashir* (Pre-Trial) ICC-02/05-01/09.

¹⁰ Ervin Staub, *Overcoming Evil: Genocide, Violent Conflict, and Terrorism* (Oxford University Press 2011).

¹¹ Stephen Reyna, 'The Disasters of War in Darfur' (2011) 31 *Third World Quarterly* 1297.

¹² Robert Merton 'Social Structure and Anomie' (1938) 3 *American Sociological Review* 672.

¹³ Janie Leatherman, *Sexual Violence and Armed Conflict* (Polity 2011).

¹⁴ Ervin Staub, *Overcoming Evil: Genocide, Violent Conflict, and Terrorism* (Oxford University Press 2011).

There are three sets of cultural preconditions present within Darfur which are particularly relevant to the onset of sexual violence: i) Tension between the African and Arab population,¹⁵ ii) The patriarchal nature of Sudanese culture,¹⁶ and iii) Hegemonic masculinities.¹⁷ There was a long history of harmony between the Arab and African populations. Under a sultanate in Darfur, the native Fur population happily adopted Islam. Some became bilingual and accepted Arabs as equals.¹⁸ It is a more recent history in which there were disputes over land and water access which exacerbated the tension between the two groups.¹⁹ Government policy contributed to the construction of tensions by favouring the Arab population and thus creating a dichotomy between the two groups.²⁰ These tensions arose concurrently with divisive ideology which will be discussed further when considering meso levels of analysis. Despite the apparent growing resentment between the African and Arab populations, Hale suggests that the racial dichotomy argument is over simplified, and instead the symbolic importance of women within the culture is more significant in explaining sexual violence against women in Darfur.²¹

Zurbriggen argues that sexual violence in conflict is more likely in patriarchal cultures because it has a greater effect, particularly where patriarchy influences a culture of honour and shame and therefore acts as an efficient weapon of war.²² The Arab

¹⁵ John Hagan, Wenona Rymond-Richmond and Patricia Parker 'The Criminology of Genocide: The Death and Rape of Darfur' (2005) 43 *Criminology* 525; Stephen Reyna, 'The Disasters of War in Darfur' (2011) 31 *Third World Quarterly* 1297.

¹⁶ Sondra Hale, 'Rape as a Marker and Eraser of Difference: Darfur and the Nuba Mountains' in Laura Sjoberg and Sandra Via (eds), *Gender, War and Militarism: Feminist Perspectives* (Praeger 2010).

¹⁷ Karin Willemsse, 'The Darfur War: Masculinity and the Construction of a Sudanese National Identity' in Salah Hassan and Carina Ray (eds), *Darfur and the Crisis of Governance in Sudan: A Critical Reader* (Cornell University Press 2009).

¹⁸ Julie Flint and Alex de Waal, *Darfur: A New History of a Long War* (Zed Books 2005) 3

¹⁹ John Hagan and Wenona Rymond-Richmond, *Darfur and the Crime of Genocide* (Cambridge University Press 2009).

²⁰ *ibid.*

²¹ Sondra Hale, 'Rape as a Marker and Eraser of Difference: Darfur and the Nuba Mountains' in Laura Sjoberg and Sandra Via (eds), *Gender, War and Militarism: Feminist Perspectives* (Praeger 2010).

²² Eileen Zurbriggen 'Rape, War and the Socialisation of Masculinity: Why Our Refusal to Give Up War Ensures that Rape Cannot Be Eradicated' (2010) 34 *Psychology of Women Quarterly* 538.

influence in Sudan has heightened this patriarchy, and in Arab culture, the males of the family also have their power reasserted in the concept of kinship which creates and reinforces patriarchal lineage.²³ In contexts such as Darfur, the cultural significance of the woman enables males to maintain power.²⁴ This occurred in Darfur following Islamisation in the nineteenth century, a period where tighter restrictions began to be imposed upon women.²⁵ During this period of Islamisation, it was the women who maintained pre-Islamic tradition, passing it down between generations.²⁶ This is indicative that women, although still symbolic of the culture, previously had more power and agency. This contributes to the sexual violence of women in conflict. It is understood that where the woman is symbolic of the community or carries the honour of the family, rape has a greater impact because it is an attack on the body, the family, and the community. Thus, it is an effective weapon of war in contexts where patriarchy is particularly prevalent.²⁷ While the men and boys are killed, women are raped as a method of deracination and humiliation. This also defaces the culture of the outgroup,²⁸ signifying gendered victimisation impacted upon by patriarchy. Without the patriarchal beliefs about the power and oppression of women, sexual violence would not be so prevalent.

In addition to the influence of patriarchy is that of hegemonic masculinities. Drawing from Gramsci's concept of cultural hegemony,²⁹ Connell considered concepts of power, patriarchy and oppression, and suggested that masculinities are

²³ Margot Badran 'Islam, Patriarchy and Feminism in the Middle East' (1986) 4 Trends in History 49; Suad Joseph 'Patriarchy and Development in the Arab World' (1996) 4 Gender and Development 14.

²⁴ Sondra Hale, 'Rape as a Marker and Eraser of Difference: Darfur and the Nuba Mountains' in Laura Sjoberg and Sandra Via (eds), *Gender, War and Militarism: Feminist Perspectives* (Praeger 2010).

²⁵ Margot Badran 'Islam, Patriarchy and Feminism in the Middle East' (1986) 4 Trends in History 49.

²⁶ Rex O'Fahey, *The Darfur Sultanate: A History* (HURST 2008) 38.

²⁷ Sara Meger 'Rape of the Congo: Understanding Sexual Violence in the Democratic Republic of Congo' (2010) 28 Journal of Contemporary African Studies 119.

²⁸ Sondra Hale, 'Rape as a Marker and Eraser of Difference: Darfur and the Nuba Mountains' in Laura Sjoberg and Sandra Via (eds), *Gender, War and Militarism: Feminist Perspectives* (Praeger 2010).

²⁹ Antonio Gramsci, *Selections from the Prison Notebooks* (Lawrence and Wishart 1971).

practiced in a variety of ways within different contexts.³⁰ Hegemonic masculinities draw on the social construction of gender to include heterosexuality, power and authority.³¹ Messerschmidt applied the concept of hegemonic masculinities to crime in his structured action theory, suggesting that when men are not able to meet ideals of masculinity through channels such as the labour market, they will commit crime in order to attain these ideals.³² In Darfur, in addition to traditional hegemonic masculinities of providing for the family and virility, these include the ideals of control of a woman and are also combined with the attainment of a Sudanese Muslim identity.³³ Hegemonic masculinities have previously been discussed in the context of rape in the Rwandan genocide where with an absence of land and income, men were unable to marry and thus achieve hegemonic masculine ideals.³⁴ Likewise, in Darfur, marriage and therefore the control of a woman is a socially constructed ideal for men. In addition, following the drought and with growing competition for land, men found it increasingly difficult to meet hegemonic masculine ideals of being able to marry and provide for their family. If structured action theory applies in this context, it follows that in these conditions men will commit crime in order to achieve these ideals. In Darfur, this resulted in looting and the expulsion of Africans from their land which would enable the Janjaweed to achieve economic ideals and the rape of women in order to achieve the power and control over women. However, Jefferson argues that the theory is over reliant on socialisation, and greater emphasis needs to be placed on the individual life course and psychosocial understandings of masculinity, and therefore the impact of individual masculinities will be discussed here when considering micro levels of analysis.³⁵

³⁰ Raewyn Connell, *Masculinities* (Polity 1995).

³¹ *ibid.*

³² James Messerschmidt, *Crime as Structured Action: Gender, Race, Class and Crime in the Making* (SAGE 1997).

³³ Karin Willemse, 'The Darfur War: Masculinity and the Construction of a Sudanese National Identity' in Salah Hassan and Carina Ray (eds), *Darfur and the Crisis of Governance in Sudan: A Critical Reader* (Cornell University Press 2009).

³⁴ Miranda Alison 'Wartime Sexual Violence: Women's Human Rights and Questions of Masculinity' (2007) 33 *Review of International Studies* 75.

³⁵ Tony Jefferson 'Subordinating Hegemonic Masculinity' (2002) 6 *Theoretical Criminology* 63.

Whilst these macro levels of analysis form a valid argument in explanation of the sexual violence as part of the Darfur conflict, there are further aspects of this context of mass violence to consider. The process of colonisation and decolonisation impacted greatly on Darfur and meso level of analysis factors such as ideology. While the British and the postcolonial Arab-dominated Sudanese government installed by the colonisers invested in the north of Sudan, the Darfur region was neglected, which ultimately resulted in the poverty discussed earlier. The neglect of the region led to Arab farmers encroaching on the land of their African counterparts.³⁶ This, whilst increasing tensions, also contributed to the 'us and them' discourse, positing the African population as the outgroup.³⁷ This was further compounded by the Arab supremacist ideology influenced by Libya that was encouraged by Khartoum.³⁸ This ideology was the catalyst for the Janjaweed being sanctioned and armed by the Sudanese and Libyan governments.³⁹

Meso-level factors include organisational aspects of the committal of atrocities. In the context of Darfur, this includes the arming and engagement of the Janjaweed militia by the Sudanese government.⁴⁰ The use of pro-government militia provides efficiency as they are cheaper, are not trained by the state, and have valuable local knowledge.⁴¹ Furthermore, the use of militia gives the government deniability, as they can avoid accountability.⁴² However, in this context, whilst the state has repeatedly publicly

³⁶ John Hagan, Wenona Rymond-Richmond and Patricia Parker 'The Criminology of Genocide: The Death and Rape of Darfur' (2005) 43 *Criminology* 525.

³⁷ John Hagan, Wenona Rymond-Richmond and Alberto Palloni, 'Racial Targeting of Sexual Violence in Darfur' (2009) 99 *American Journal of Public Health* 1386.

³⁸ Julie Flint, 'The Other Conflict: Inter-Arab Conflict in Darfur' (2010) Small Arms Survey Working Paper 22 <<http://www.smallarmssurveysudan.org/fileadmin/docs/working-papers/HSBA-WP-22-The-Other-War-Inter-Arab-Conflict-in-Darfur.pdf>> accessed 7 May 2016.

³⁹ Amy Eckert, 'The Cosmopolitan Test: Universal Morality and the Challenge of the Darfur Genocide' in Steven Roach (ed), *Governance, Order and the International Criminal Court* (Oxford University Press 2009).

⁴⁰ John Hagan and Wenona Rymond-Richmond, *Darfur and the Crime of Genocide* (Cambridge University Press 2009).

⁴¹ Sabine Carey, Michael Colaresi and Neil Mitchell, 'Governments, Informal Links to Militias, and Accountability' (2015) 59 *Journal of Conflict Resolution* 850.

⁴² Paul Staniland, 'Between a Rock and a Hard Place: Insurgent Fratricide, Ethnic Defection, and the Rise of Pro-State Paramilitaries' (2012) 56 *Journal of Conflict Resolution* 16.

denied knowledge of sexual violence in Darfur,⁴³ they have been complicit in the attacks. Sexual violence was increased when government forces attacked with the Janjaweed,⁴⁴ and a recent report has cited government forces perpetrating rape alone.⁴⁵ While initially, the Janjaweed may have been used to avoid accountability, since evidence of the government's involvement in crimes breaching international law and the unenforced arrest warrants against government personnel, the government have continued to act alone with impunity.⁴⁶ Further evidence from the Small Arms Survey suggests that while the Janjaweed militia initially served the government, the lack of fulfilment of rewards offered by the government led to the lack of their control over the Janjaweed.⁴⁷ This may have contributed to the changes seen in Darfur, with the Sudanese government forces acting alone in the perpetration of sexual violence in recent years.⁴⁸

Other social structures, such as the police, have also compounded the victimisation of women. Women fail to report being victims of sexual violence due to fear of being arrested or facing further violence at the hands of police and other government officials.⁴⁹ There have been reports of people being arrested for speaking publicly about attacks perpetrated by government forces.⁵⁰

⁴³ John Hagan and Wenona Rymond-Richmond, *Darfur and the Crime of Genocide* (Cambridge University Press 2009).

⁴⁴ John Hagan, Wenona Rymond-Richmond and Alberto Palloni, 'Racial Targeting of Sexual Violence in Darfur' (2009) 99 *American Journal of Public Health* 1386.

⁴⁵ Human Rights Watch, 'Mass Rape in North Darfur: Sudanese Army Attacks Against Civilians in Tabit' (*Human Rights Watch*, 11 February 2015) <<https://www.hrw.org/report/2015/02/11/mass-rape-north-darfur/sudanese-army-attacks-against-civilians-tabit#page>> accessed 5 May 2016.

⁴⁶ Human Rights Watch, 'Mass Rape in North Darfur: Sudanese Army Attacks Against Civilians in Tabit' (*Human Rights Watch*, 11 February 2015) <<https://www.hrw.org/report/2015/02/11/mass-rape-north-darfur/sudanese-army-attacks-against-civilians-tabit#page>> accessed 5 May 2016.

⁴⁷ Julie Flint, 'Beyond Janjaweed: Understanding the Militias of Darfur' (2009) Small Arms Survey Working Paper 17 <<http://www.smallarmssurveysudan.org/fileadmin/docs/working-papers/HSBA-WP-17-Beyond-Janjaweed.pdf>> accessed 10 May 2016.

⁴⁸ Human Rights Watch, 'Mass Rape in North Darfur: Sudanese Army Attacks Against Civilians in Tabit' (*Human Rights Watch*, 11 February 2015) <<https://www.hrw.org/report/2015/02/11/mass-rape-north-darfur/sudanese-army-attacks-against-civilians-tabit#page>> accessed 5 May 2016.

⁴⁹ *ibid.*

⁵⁰ *ibid.*

In addition, where female victims of sexual violence are unable to prove they have been raped, until recently they faced arrest under Zina laws which govern sexual conduct.⁵¹ This dual victimisation for women also prevented them obtaining medical care due to fear of further harm and arrest from other social structures.⁵² It was only in 2015, following pressure from women's activists, that rape was partially disassociated from Zina.⁵³ It is clear that while the militia and government forces are the primary perpetrators of sexual violence, other social structures also impact upon the problem. As long as women fear reporting their experiences of this type of victimisation, the less their perpetrators face consequences for their actions, which enables them to re-offend with impunity.

Alvarez argues that ideology is crucial in the perpetration of genocidal crimes because it creates a shared set of beliefs and values which allows the destruction of a population.⁵⁴ Perpetrators of genocide accept this ideology which justifies them acting in a way in which they would not act during peacetime. Alvarez suggests that ideology springs from both rational and irrational reasoning, and can include notions of nationalism, past victimisation and dehumanisation, all of which contribute to the justification of genocidal violence.⁵⁵ Arab supremacist ideology in Darfur included all these factors. Past victimisation of land conflicts and attacks from rebel African groups in the region provided logical reasoning for the promotion of the African group as a threat;⁵⁶ the influence of Libya and the Arab-dominated Sudanese government created a

⁵¹ Kelly Askin, 'Prosecuting Gender Crimes Committed in Darfur' in Samuel Totten and Eric Markusen (eds), *Genocide in Darfur* (Routledge 2006).

⁵² Human Rights Watch, 'Mass Rape in North Darfur: Sudanese Army Attacks Against Civilians in Tabit' (*Human Rights Watch*, 11 February 2015) <<https://www.hrw.org/report/2015/02/11/mass-rape-north-darfur/sudanese-army-attacks-against-civilians-tabit#page>> accessed 5 May 2016.

⁵³ Liv Tonneson and Samia al-Nagar, 'Women and Girls Caught Between Rape and Adultery in Sudan: Criminal Law Reform 2005-2015' (2015) CMI Report R 2015:10 <<https://www.cmi.no/publications/file/5661-women-and-girls-caught-between-rape-and-adultery.pdf>> accessed 6 May 2016.

⁵⁴ Alex Alvarez, *Genocidal Crimes* (Routledge 2010).

⁵⁵ *ibid.*

⁵⁶ John Hagan and Wenona Rymond-Richmond, *Darfur and the Crime of Genocide* (Cambridge University Press 2009).

sense of Arab identity and the creation of an Arab nation;⁵⁷ and dehumanising ideology including the ‘us’ and ‘them’ rhetoric which infers a less human ‘them’. In raping a woman, the outgroup is terrorised, thereby weakening the spirit of the outgroup. In impregnating her, the Arab nation is being strengthened, while the outgroup is weakened and the population reduced, particularly since women who have been raped in the region rarely remarry.⁵⁸ Finally, in dehumanising the victims of sexual violence, perpetrators are better able to justify their actions. This has been evidenced in Hagan et al. who found that incidence of sexual violence was increased where racial epithets were heard.⁵⁹

All of these factors are particularly important with regards to rape and suggested that ideological propaganda was an influential factor in the perpetration of sexual violence.⁶⁰ Although Stiglmeier’s research was in Bosnia, propaganda was widely used by state media in Sudan.⁶¹ The term Tora Bora, taken from the name of a place linked to al-Qaeda, was widely used by state representatives and throughout state media to depict the African rebels as terrorists.⁶² It was also used in conjunction with an American wild west analogy, depicting the African rebel groups as highway men who would steal from the Arabs. This propaganda provided the Sudanese government with justification to use the Janjaweed militia in fighting against the African population. This was a further method of dehumanisation and to spread the ideology of hatred towards the African population efficiently. It provides further evidence that the Arab/African divide was politically constructed.

⁵⁷ Julie Flint, ‘The Other Conflict: Inter-Arab Conflict in Darfur’ (2010) Small Arms Survey Working Paper 22 <<http://www.smallarmssurveysudan.org/fileadmin/docs/working-papers/HSBA-WP-22-The-Other-War-Inter-Arab-Conflict-in-Darfur.pdf>> accessed 7 May 2016.

⁵⁸ John Hagan and Wenona Rymond-Richmond, *Darfur and the Crime of Genocide* (Cambridge University Press 2009).

⁵⁹ John Hagan, Wenona Rymond-Richmond and Alberto Palloni, ‘Racial Targeting of Sexual Violence in Darfur’ (2009) 99 American Journal of Public Health 1386.

⁶⁰ Alexandra Stiglmeier, ‘The Rapes in Bosnia-Herzegovina’ in Alexandra Stiglmeier (ed), *Mass Rape: The War Against Women in Bosnia-Herzegovina* (University of Nebraska Press 1993).

⁶¹ John Hagan and Wenona Rymond-Richmond, *Darfur and the Crime of Genocide* (Cambridge University Press 2009).

⁶² *ibid.*

Thus far, this essay has discussed the macro and meso levels of analysis in explaining sexual violence in Darfur. Whilst they explain the wider context of the violence and how the state participates, this work would not be complete without a discussion of meso level factors such as group processes and individual factors such as denial. Sykes and Matza's techniques of neutralisation can be applied to mass sexual violence in this context.⁶³ At some point between adopting the aforementioned Arab supremacist ideology and committing an offence of sexual violence, an individual must go through an internal process in order to neutralise their actions. Cohen applied neutralisation theory to state crime, and suggested that denial of responsibility could include authorisation and dehumanising the victims.⁶⁴ Sexual victimisation in Darfur is more likely as part of a joint attack from both government forces and Janjaweed militia,⁶⁵ thus providing denial of responsibility for perpetrators. Where the actions are authorized by a legitimate body such as the government, one can deny any wrongdoing. Similarly, women were more likely to be victims of sexual violence where racial epithets and dehumanising language was heard.⁶⁶ In dehumanising an individual or group, the perpetrator eliminates the normal human moral obligation to them and therefore denies wrongdoing when committing an act of sexual violence against them.⁶⁷

Whilst techniques of neutralisation can go some way in explaining how an individual can commit these kinds of atrocities, not all perpetrators are in denial of their actions. Stiglmeier found in her research on mass rape in Bosnia that some perpetrators did feel guilt and remorse after committing acts of sexual violence, and knew it was wrong while they were committing the offence.⁶⁸ They

⁶³ Gresham Sykes and David Matza, 'Techniques of Neutralisation' (1957) 22 *American Sociological Review* 664.

⁶⁴ Stanley Cohen, *States of Denial* (Polity Press 2002).

⁶⁵ John Hagan, Wenona Rymond-Richmond and Alberto Palloni, 'Racial Targeting of Sexual Violence in Darfur' (2009) 99 *American Journal of Public Health* 1386.

⁶⁶ *ibid.*

⁶⁷ Stanley Cohen, *States of Denial* (Polity Press 2002).

⁶⁸ Alexandra Stiglmeier, 'The Rapes in Bosnia-Herzegovina' in Alexandra Stiglmeier (ed), *Mass Rape: The War Against Women in Bosnia-Herzegovina* (University of Nebraska Press 1993).

chose to keep quiet and to continue in their actions in order to live up to expectations of masculinity. In cases where the perpetrators were not physically aroused or did not appear to enjoy it, they were mocked.⁶⁹ Meger also found that masculinities impacted upon the perpetration of sexual violence and suggested that the act of rape fulfilled the ideal of the sexually potent male fighter.⁷⁰ Moreover, military socialisation intensifies these masculinities and therefore increases the propensity to commit sexual violence.⁷¹ If an individual is socialised to perpetrate acts of violence towards the enemy, it is therefore unsurprising that acts of sexual violence are perpetrated against female enemies. Although these studies explain the impact of masculinities in Bosnia and the Congo, it is not known if they are applicable to the sexual violence being perpetrated in Darfur. Sexual violence differs between contexts in extent. For example, who is victimised and how they are victimised, who perpetrates the violence and why, varies.⁷² Given this variance, it is not possible to generalise micro level theories, and therefore it is important for qualitative research to be undertaken in Darfur in order to gain a complete understanding of micro levels of analysis.

A further criminological theory that can be applied to this context of mass violence is routine activities theory. Cohen and Felson suggested that offending will occur where there is a motivated offender, a suitable target, and the absence of a capable guardian.⁷³ This theory can be applied to Darfur as there are many motivated offenders, and suitable targets are in abundance as a majority of those killed in the genocide were fighting aged men, leaving women and children.⁷⁴ The reduced male population also

⁶⁹ Alexandra Stiglmeier, 'The Rapes in Bosnia-Herzegovina' in Alexandra Stiglmeier (ed), *Mass Rape: The War Against Women in Bosnia-Herzegovina* (University of Nebraska Press 1993).

⁷⁰ Sara Meger 'Rape of the Congo: Understanding Sexual Violence in the Democratic Republic of Congo' (2010) 28 *Journal of Contemporary African Studies* 119.

⁷¹ Eileen Zurbriggen 'Rape, War and the Socialisation of Masculinity: Why Our Refusal to Give Up War Ensures that Rape Cannot Be Eradicated' (2010) 34 *Psychology of Women Quarterly* 538.

⁷² Elisabeth Wood 'Variation in Sexual Violence During War' (2006) 34 *Politics and Society* 307.

⁷³ Lawrence Cohen and Marcus Felson 'Social Change and Crime Rate Trends: A Routine Activity Approach' (1979) 44 *American Sociological Review* 588.

⁷⁴ John Hagan and Wenona Rymond-Richmond, *Darfur and the Crime of Genocide* (Cambridge University Press 2009).

created the absence of a capable guardian. Meger echoed this theory, arguing that rape in conflict occurs because there are fewer or weaker structures of social control in times of conflict which also contributes to the absence of a capable guardian, and that perpetrators have always been capable of sexual violence in peacetime but that the conflict provides increased opportunity.⁷⁵ Although Meger's research was undertaken in the Congo, it can be applied to this context.⁷⁶ In the Darfur region, social structures were already weak due to the lack of investment in infrastructure in the region and were further weakened in the conflict.⁷⁷

Finally, given that there were two groups committing acts of sexual violence – Sudanese government forces and the Janjaweed - a further micro-level factor to consider here is group cohesiveness. When there is solidarity within a group, there is less tension within a group which is then directed externally to the outgroup, in this context the African population⁷⁸. Hogg and Abrams further suggested that individuals are attracted to a group where they are rewarded and where they are dependent on one another to meet their needs.⁷⁹ In times of conflict, sexual violence is a reward often encouraged or turned a blind eye to by senior officers in order to increase morale.⁸⁰ Furthermore, sexual violence also occurred most frequently in Darfur during joint attacks by Sudanese government forces and the Janjaweed militia.⁸¹ Therefore, it could be argued that the dependency between the two groups in complicity of committing genocide and the following denial of the offences created cohesiveness, which then allowed for the acceptability and permissiveness of mass sexual violence. Group cohesiveness is important in the explanation of sexual violence in conflict as gang

⁷⁵ Sara Meger 'Rape of the Congo: Understanding Sexual Violence in the Democratic Republic of Congo' (2010) 28 *Journal of Contemporary African Studies* 119.

⁷⁶ *ibid.*

⁷⁷ John Hagan and Wenona Rymond-Richmond, *Darfur and the Crime of Genocide* (Cambridge University Press 2009).

⁷⁸ Michael Hogg and Dominic Abrams, *Social Identifications* (Routledge 1988) 95.

⁷⁹ *ibid* 96.

⁸⁰ Alexandra Stiglmeier, 'The Rapes in Bosnia-Herzegovina' in Alexandra Stiglmeier (ed), *Mass Rape: The War Against Women in Bosnia-Herzegovina* (University of Nebraska Press 1993).

⁸¹ John Hagan and Wenona Rymond-Richmond, *Darfur and the Crime of Genocide* (Cambridge University Press 2009).

rapes are particularly prevalent in these circumstances, creating loyalty within a group, and in the case of Darfur, between the Sudanese government forces and the Janjaweed.⁸²

By analysing a selection of factors of macro, meso and micro levels of analysis, it is clear that each level provides some explanation of the problem of sexual violence in the conflict in Darfur. However, given the complexity of the problem, it is helpful to consider all three levels. The macro level factors of difficult life conditions and cultural preconditions could provide warning signs for future conflicts. Meso level factors such as ideology and bureaucratic organisation provide justification and a catalyst for action, and there are many micro explanations to explain why individuals and groups behave in this way. Whilst these micro explanations provide convincing arguments, most individuals would not commit acts of sexual violence if the macro and meso level factors analysed in this essay were not also present. Therefore, it is clear that no single level of an analysis can completely explain a phenomenon. It is therefore important to explore the interplay between multiple levels of analysis in order to gain a broader understanding. It has also been shown in this work that mainstream criminological theory, such as Cohen and Felson's routine activities theory and Sykes and Matza's techniques of neutralisation, ordinarily used to explain domestic street crime, can be applied to genocide and international crimes including mass sexual violence in the context of war.⁸³ This essay has discussed literature from criminological theoretical frameworks and research on various contexts of mass sexual violence, and whilst some attempt has been made to apply their theories to the context of Darfur, each context is different, and further research in the Darfur region, considering multiple levels of analysis, is needed to gain a thorough understanding of both victimisation and perpetration.⁸⁴

⁸² Miranda Alison 'Wartime Sexual Violence: Women's Human Rights and Questions of Masculinity' (2007) 33 *Review of International Studies* 87.

⁸³ Lawrence Cohen and Marcus Felson 'Social Change and Crime Rate Trends: A Routine Activity Approach' (1979) 44 *American Sociological Review* 588; Gresham Sykes and David Matza, 'Techniques of Neutralisation' (1957) 22 *American Sociological Review* 664.

⁸⁴ Elisabeth Wood 'Variation in Sexual Violence During War' (2006) 34 *Politics and Society* 307.

These are just a few ways of explaining the problem in relation to various levels of analysis, however, the situation in Darfur is far too complex to explore all the contributing factors within this limited space.

Opting out of the CISG: is the law too complicated for such a popular choice?

*Océane Girard*⁸⁵

The United Nations Convention on Contracts for the International Sale of Goods (CISG) provides for the possibility to derogate from its provisions. Traders often choose not to apply the convention. This exclusion raises many questions, the first one of these questions simply being “why?”. When one takes a closer look at the CISG, one may not contemplate that opting out would cause so many unexpected consequences. Opting out actually seems easy at first glance. However, it brings many practical difficulties. Thus, is the law too complicated for such a popular choice? This paper aims at demonstrating that the CISG’s opt-out article is not complicated, but rather, incomplete. Indeed, the many issues raised by the opt-out choice (such as battle over the applicable law or effectiveness of the opt-out clause) may make traders reconsider their willingness to opt out of the CISG. As this paper underlines, this is due to the lack of precision on the opt-out option in the CISG. This paper will firstly show that opting out is a popular choice and give the reasons for such popularity before examining the ways to opt out. Secondly, this paper will draw a conclusion on the so-called complication of the opt-out solution. Finally, it will explain the difficulties that opting out raises and attempt to provide solutions. Such solutions would, for instance, rely on international cooperation or improving the CISG itself.

The United Nations Convention on Contracts for the International Sale of Goods (CISG) was created to facilitate the relationships between traders. The CISG seeks to reach a noble goal: the harmonisation of sale of goods law. As the preamble of the convention states, the unification of this area of law would ‘contribute to the removal of legal barriers in international trade and promote the development of international trade’.⁸⁶ The CISG has been adopted by 85 countries. In these countries, the CISG has become part of the national law. Thus, provided that the conditions of application are fulfilled, the CISG applies automatically.

⁸⁵ LL.M. Candidate, The University of Manchester, School of Law.

⁸⁶ United Nations Convention on Contracts for the International Sale of Goods (adopted 11 April 1980, entered into force 11 January 1988) (CISG) Preamble.

According to Article 1, the CISG applies to contracts of sale of goods between traders whose places of business are in different contracting states or where, due to the rules of international private law, the law of a contracting state is deemed to be applicable.⁸⁷

However, the CISG also provides the possibility for traders to exclude its application.⁸⁸ Article 6 constitutes an opt-out opportunity. Therefore, the absence of the opt-out clause is a 'negative applicability requirement'.⁸⁹ It seems that this article is often used by traders; several surveys⁹⁰ have shown that most transactions are made outside the CISG. This is paradoxical. Since the convention has been adopted by so many countries, its success appears incontestable. Yet, when one takes a closer look at the practice, the CISG is not as often applied as it should be. The popularity of opt-out provisions contrasts with its alleged large scope of application. What looks even more puzzling is that the wording of Article 6 is so concise that no indications on how to opt out are provided. Article 6 provides that '[t]he parties may exclude the application of this Convention or, subject to article 12, derogate from or vary the effect of any of its provisions.'⁹¹ Consequently, traders are left with no instructions as how to exclude the application of the CISG. This can lead to unwanted results. The parties may have thought they have opted out of the convention but when a dispute arises and goes before a judge, the court finds that the convention is still applicable due to a defect in the wording of the opt-out clause. As there are few guidelines on the way to opt out, parties can be misled.

⁸⁷ CISG art 1.

⁸⁸ CISG art 6.

⁸⁹ Franco Ferrari, "'Forum Shopping" despite International Uniform Contract Law Conventions' (2002) 51 ICLQ 701.

⁹⁰ Peter L. Fitzgerald, 'The International Contracting Practices Survey Project: an Empirical Study of the Value and Utility of the United Nations' Convention on the International Sale of Goods (CISG) and the Unidroit Principles of International Commercial Contracts to Practitioners, Jurists, and Legal Academics in the United States' (2008) 27 JL&Com 1; Martin F. Koehler and Guo Yujun, 'The Acceptance of the Unified Sales Law (CISG) in Different Legal Systems' (Spring 2008) 20 PaceIntlLRev 45; Ingeborg Schwenzer and Christopher Kee, 'International Sales Law - The Actual Practice' (2011) 29 PennStIntlLRev 425; Luiz Gustavo Meira Moser, 'Parties' Preference in International Sales Contracts: an Empirical Analysis of the Choice of Law' (2015) 20 RevDrUnif 19.

⁹¹ CISG (n 88).

This vagueness of the law begs the question as to whether it is too complicated to opt out of the CISG? There are obvious reasons to support that a problem appears in the words of Article 6. Notwithstanding these reasons, one must not deduce that the law is too complicated. Rather, it is more relevant to conclude a lack of law than a complication within. That is what this paper will try to demonstrate. First, it will underline the popularity of the opt-out choice and explain why it is so well-received. Then, the different possibilities to opt out will be examined in order to find out whether the law requires clarification. Finally, the problems raised by opting out will be explained and their potential solutions will be suggested.

I. Popularity of opting out: A mirror image of the unpopularity of the CISG

The CISG intends to favour international trade by providing for common rules. Thus exchanges between traders would benefit from such harmonised rules: if traders know what to expect from the contracts they enter into, they are likely to be less reluctant to conclude them. Harmonisation would increase trust between contracting parties and consequently develop international commerce. This is the main reason why the convention should be used in the context of international sale of goods. The CISG is also a neutral set of rules. This is another factor which could increase trading opportunities: it is easier to convince a contracting party to have the contract governed by such type of rules than by the home law of a party. However, as the position of the United-Kingdom on the CISG expresses, the convention may sometimes be seen as an inconvenient tool for international transactions. Some consider that the CISG has more drawbacks than advantages, which may add depth to why the parties often decide to opt out.⁹² This section will highlight the current trend in the (non-) application of the CISG and

⁹² Angelo Forte, 'The United Nations Convention on Contracts for the International Sale of Goods: Reason or Unreason in the United-Kingdom' [1997] UBaltLRev 51; Robert G. Lee, 'The UN Convention on Contracts for the International sale of Goods: OK for the UK?' [1993] JBL 131; Sally Moss, 'Why the United Kingdom has not ratified the CISG' [2005-06] JL&Com 483.

define the reasons why the convention is so often omitted by traders in their daily transactions.

Several surveys have shed light on the popularity of opting out. Schwenzer and Kee carried out a survey in 2009 in 85 countries.⁹³ Their results are the following:

‘Of the lawyers from CISG member states who answered this question, only 13% reported always excluding the CISG, and a further 32% reported they sometimes did so. A considerable 55% answered that they never or rarely excluded the CISG. [...] Amongst [responses from non-contracting states], 19% always excluded the CISG and 36% did so sometimes, while 45% rarely or never excluded it.’⁹⁴

Moser also produced his own survey which was conducted in 2014 in 93 jurisdictions.⁹⁵ 33.34% of the respondents indicated that they had already opted-out of the convention. The figures are even higher in Koehler and Yujun’s survey. They focused their survey on the USA, Germany and China in 2004, 2005 and 2007.⁹⁶ They found out that less than 10% of the respondents never exclude the CISG, while about 65% principally or preponderantly exclude its application. Similar results can be found in Fitzgerald’s survey.⁹⁷ Fitzgerald studied the value and utility of the CISG. His survey of the practices in the USA between 2006 and 2007 revealed that the majority of the US practitioners opt out of the CISG (55%).⁹⁸

To sum up the results of these surveys, it is safe to say that about half of the contracts involving international sales of goods exclude the CISG. For a convention whose merits have been so often

⁹³ Schwenzer and Kee, (n 90).

⁹⁴ *ibid* 434.

⁹⁵ Moser (n 90) 19.

⁹⁶ Koehler and Yujun (n 90) 45.

⁹⁷ Fitzgerald (n 90).

⁹⁸ *ibid* 14.

praised in the literature, it seems a high proportion.⁹⁹ Thus, opting out appears to be a popular choice amongst the international trade community. Traders do not seem to believe in harmonisation. As a consequence, the CISG fails to reach its main goal. This clearly indicates a deficiency in the drafting of the law.

Moreover, apart from such surveys, it is worth mentioning that the exclusion of the CISG is a well-known practice in professional organisations. International trade associations such as the Grain and Feed Trade Association, and the Federation of Oils, Seeds and Fats Associations, choose to exclude the convention from their standard terms as it is insufficiently connected with their realm of business. Moreover, some examples of sample contracts facilitate the exclusion of the convention by providing a choice for the choice of law clause (application of the CISG or opt out).¹⁰⁰ Thus the parties simply have to select their favourite position. The troubles are avoided as the exclusion clause is ready to be used. This easiness may be a factor contributing to the exclusion of the CISG.

However, the main interest brought by these surveys consists in potential reasons to explain the popularity of opt-out.

The first element of explanation that comes to mind when the question of the application of the CISG is raised is the content of the convention itself. If the parties think the convention is not in their best interest, then they will exclude it. A balance of the advantages and drawbacks offered by the CISG must be realised before taking a decision. In Koehler's survey, 35% of the respondents said they exclude the convention because they see no advantages in it, and 38% because they have no need to apply it.¹⁰¹ Thus more than a third of the respondents doubt the utility of the convention. Further, only 7.4% of the respondents of the same

⁹⁹ RM Lavers, 'CISG: To use or not to use?' (1993) 21 *IntlBusL* 10, 13; Susanne Cook, 'CISG: From the Perspective of the Practitioner' (1997-1998) 17 *J.L. & Com.* 349; Filip De Ly, 'The Relevance of the Vienna Convention for International Sales Contracts-Should We Stop Contracting Out?' [2003] *BLI* 241.

¹⁰⁰ Allison E. Butler, *A Practical Guide to the CISG: Negotiations through Litigation* (Supplement 2, Aspen Publishers 2007).

¹⁰¹ Martin F. Koehler, 'Survey Regarding the Relevance of the United Nations Convention for the International Sale of Goods (CISG) in Legal Practice and the Exclusion of its Application' (Oct 2006) <<http://www.cisg.law.pace.edu/cisg/biblio/koehler.html>> accessed 20 March 2016.

survey find legal advantages in the CISG.¹⁰² According to some, it is preferable to exclude the convention because of its uncertainty. As the convention is ‘still in process of earning recognition among courts, practitioners and merchants’, its rules may be uncertain.¹⁰³ Many concepts have no precise definition, such as fundamental breach or good faith. Moreover, the convention does not cover every aspect of the transaction. Questions about property, for example, are excluded from its scope (Article 4).

Thus, the CISG is an incomplete instrument. Additional rules are necessary to fill in the gaps. However, it is not clear which rules would apply. National law or Incoterm, for example, provide rules that could potentially be applied. It would be easier to apply these rules clearly and directly to avoid uncertainty. Many uncertainties arise from the convention. As the parties seek to escape from unpredictable results (e.g. committing a fundamental breach without being aware of it because of the lack of definition in the CISG), they will tend to exclude the application of the convention. The idea of uncertainty is reinforced by the fact that there is no uniformity in its application. Courts from different jurisdiction may reach opposite judgments on the same issue, contributing to the general expansion of uncertainty. Several authors point to the lack of case law¹⁰⁴ as a reason in favour of opting out. Due to the few cases available (although the situation is improving with more and more decisions translated and provided by the CISG database), the outcomes of conflicts between parties are not as predictable as those dealt with under domestic law, which benefits from a larger pool of precedents. Moreover, as Lavers states, many cases do not go before a judge but are settled via arbitration.¹⁰⁵ Therefore, they are not published and cannot constitute reference points for future litigants.

¹⁰² *ibid.*

¹⁰³ Susanne Cook, ‘CISG: From the Perspective of the Practitioner’ (1997-1998) 17 *JL&Com.* 350.

¹⁰⁴ Schwenzer and Kee (n 90) 435; Moser (n 90) 50.

¹⁰⁵ RM Lavers, ‘CISG: To Use or Not to Use?’ (1993) 21 *IntBusL* 10, 11.

However, surveys often show that the exclusion of the CISG is ‘not necessarily due to material drawbacks’.¹⁰⁶ The unfamiliarity of the actors of international trade with the convention is highlighted in every survey. Spagnolo identified two types of unfamiliarity: a lack of sufficient knowledge of the convention and its mechanisms, and a lack of awareness of the existence of the CISG.¹⁰⁷ It is true that when a party is unsure of the consequences certain provisions of the CISG could bring, it may tend to opt out and choose a law with which it is more familiar.

Continuing with the idea of unfamiliarity, it is argued that costs spent on learning how to use the CISG often lead the parties to exclude the convention. Becoming familiar with international statutes is both costly and time-consuming as traders and lawyers alike must provide an effort to understand something new. This investment may deter traders from applying the convention as it is simply easier to keep on using their usual rules.

Spagnolo also highlights that the exclusion of the application of the CISG may come from the ‘bargaining strength’¹⁰⁸ of the parties. When a party has such strength, he can urge the other party to apply his choice of law. Therefore, if the strongest party does not want the CISG to apply, the other party may have to follow that choice. In the same article, a very interesting example of this point is explained. Most lawyers in China use the CISG in international sale of goods contracts as it often seems more acceptable to the other party, the other alternative being Chinese law. Thus, Chinese traders, in order to increase their opportunities to conclude commercial deals with foreign parties, choose to rely on the CISG. As China is nowadays a powerful trading partner on the international commercial scene, it can more and more impose its conditions. Consequently, trading partners of China may have to adapt to Chinese traders’ choice and apply the convention.

¹⁰⁶ Christian Nick, ‘The Case for Ireland’s Accession to the UN Convention on Contracts for the International Sale of goods’ (2010) 32 *DublinULJ* 349.

¹⁰⁷ Lisa Spagnolo, ‘A Glimpse through the Kaleidoscope: Choices of Law and the CISG (Kaleidoscope Part I)’ (2009) 1 *VindobonaJIntlComL&Arbitration* 136.

¹⁰⁸ *ibid* 149.

Further favouring the non-application of the CISG is an idea which Spagnolo has called ‘path dependence’.¹⁰⁹ According to this expression, opt-outs may occasionally occur only as the parties use the same standard contract they have been using for years. By using a contract that is not up to date with the latest legal changes, the parties simply keep ignoring the convention. This may be due to the costs of hiring a legal adviser to draft a new version of the contract, though the behaviour of some lawyers makes the problem even worse: they themselves follow their usual practice habits and forget to take their client’s interests into consideration. Instead of taking the CISG into consideration, they simply put it aside.

Then comes the idea of group pressure. Some may be tempted to opt out of the convention because others do. For instance, if the leading company of a certain type of industry systematically opts out, its competitors may want to do the same to become closer to that company. As Spagnolo states, ‘they prefer to stick with the party line’.¹¹⁰ The influence of the network should not be underestimated in those types of situations. Thus, opting out of the CISG is not always a decision that is taken after careful consideration of its advantages and drawbacks, but rather an impulsive and irrational option.

To sum up, opting out is more than common in the trade world and is backed up by more or less logical reasons, as for instance, the lack of precision of the convention or group pressure. If it is such a common practice to opt out, it should not be very difficult to do so. Given the current status of practice, this statement may have to be reconsidered.

¹⁰⁹ Lisa Spagnolo, ‘Rats in Kaleidoscope: Rationality, Irrationality, and the Economics and Psychology of opting out of the CISG (Kaleidoscope Part II)’ (2009) 1 *VindobonaJIntlComL&Arbitration* 157.

¹¹⁰ *ibid* 166.

II. How to opt out: A guide for traders and their lawyers

Very often, the parties intend to exclude the application of the convention. However, this is more difficult than it seems to make this intention a concrete clause of their contract. The parties may not have legal knowledge or time to think about how to draft the choice of law clause. This is why the various ways to contract out will be explained in this part. After exposing the different techniques, a conclusion on the complicated character of the law will be drawn.

Article 6 is a rather short article:

“The parties may exclude the application of this Convention or, subject to article 12, derogate from or vary the effect of any of its provisions.”¹¹¹

There seems to be three requirements in order to effectively opt out.¹¹² First, the parties must know that the CISG would apply to their contract. Then, they must make it clear that they want to avoid the application of the convention. Finally, opting out cannot be unilateral and must be consented by both parties. However, these are only suppositions as drafters of the convention were not precise on how to opt out, and thus traders had to make their own interpretation.

It is to be understood from Article 6 that the first means to opt out is by including an express reference to the exclusion of the convention. This is the most common way to opt out as Koehler’s survey pointed out, with 77% of the respondents excluding the CISG by express exclusion¹¹³ (and more than 81% in the updated version of the survey including Chinese respondents¹¹⁴). For instance, the parties could draft a clause stating that ‘pursuant to Article 6 of the

¹¹¹ CISG Art 6.

¹¹² Morten M. Fogt, ‘Private International Law Issues in Opt-out and Opt-in Instruments of Harmonization: the CISG and the Proposal for a Common European Sales Law’ (2012-2013) 19 *ColumJEurL* 90.

¹¹³ Koehler (n 103).

¹¹⁴ Koehler and Yujun (n 90) 48.

CISG, the parties hereby expressly opt out of the CISG'.¹¹⁵ This would clearly underline their intent not to have the transaction governed by the convention. The reference to Article 6 in the clause is necessary because the CISG 'governs the manner of exclusion'.¹¹⁶ Indeed, this is logical as the convention applies as long as the parties have not opted out, provided that the conditions of Article 1 are fulfilled. Thus, the way to exclude its application must be made in accordance with its provisions. This type of clause could contain indications as to the applicable law or not; in the latter case the applicable law will result from the rules of international private law.

There is no time limit to opt out: the parties can decide to opt out at the time of conclusion of the contract or after that time. Thus, it gives them more freedom to conduct their transactions.

The Advisory Council insists that the intention to exclude the convention 'should be clearly manifested'.¹¹⁷ The requirement of clear intent finds its roots in the general principles of the CISG. It supports good faith and uniformity.¹¹⁸ The requirement of a clear intent is necessary because the goal of the convention is to harmonise trade transactions around the world and also, as stated above, the CISG automatically applies when its conditions of applicability are fulfilled. If no clear intention was required, it would be too easy for parties to consider that another set of rules governs their transaction and no harmonisation would take place. Besides, the automatic application requires a clear willingness to exclude it as silence does not constitute a sufficient expression of intent. This intention can be expressed in the contract itself or in general conditions as long as these conditions have been incorporated in the agreement¹¹⁹ and both parties are aware of them. Moreover, the Advisory Council added that the intention can be drawn from words

¹¹⁵ Allison E. Butler, *A Practical Guide to the CISG: Negotiations through Litigation* (Supplement 2, Aspen Publishers 2007) Form B-1.

¹¹⁶ CISG-Advisory Council, 'Opinion No. 16 Exclusion of the CISG under Article 6' [2014] <<http://www.cisg.law.pace.edu/cisg/CISG-AC-op16.html>> accessed 21 March 2016 para 2.

¹¹⁷ *ibid* para 3.

¹¹⁸ Allison E. Butler, *A Practical Guide to the CISG: Negotiations through Litigation* (Supplement 2, Aspen Publishers 2007) ch 2.

¹¹⁹ Appellate Court Jura 3 November 2004 <<http://cisgw3.law.pace.edu/cases/041103s1.html>> accessed 21 March 2016.

and/or conduct of the parties.¹²⁰ The advisory council also envisaged the case of doubt as to the intention of the parties.¹²¹ In such circumstances, judges must not infer exclusion. Therefore, if there is no clear intention to opt out, the convention will not be excluded.

Implied exclusions have brought discrepancies among courts' decisions and drafters of the convention.¹²² Article 6 does not contain any mention of such a type of opt-out although the text on which the convention is based (Article 3 ULIS) provides that the parties could impliedly opt out. Schlechtriem and Butler explain in 'UN Law on International Sales: The UN Convention on the International Sale of Goods'¹²³ that this is to impose barriers on judges so that they do not arrive at implied exclusions too easily. However, that does not mean that implied opt-outs are impossible. For example, a way to exclude the application of the convention would be not to fulfill the requirements for its application like in case of a sale of goods for personal use. The Advisory Council declared that the requirement of a clear intent to opt out is also to be applied for implied exclusions¹²⁴ and is even more important in these cases as courts have been reluctant to find opt-outs when a clear intent to exclude is lacking. To support that point, one can refer to several decisions from various jurisdictions. Courts in Austria, Switzerland, Italy, or Greece,¹²⁵ have recognised implied exclusions. In 2001, Austrian judges held that a choice of law without explicit declaration that the CISG is excluded does not constitute an implicit exclusion (because, in that particular case, the CISG was part of the chosen law). An implicit exclusion is only

¹²⁰ CISG-AC (n116) para 3.6.

¹²¹ CISG-AC (n 118) para 3.7.

¹²² United Nations Conference on Contracts for the International Sale of Goods (A/CONF.97/19 - Official Records, 1980).

¹²³ Petra Butler and Peter Schlechtriem, *UN Law on International Sales* (Springer-Verlag Berlin Heidelberg 2009) 19.

¹²⁴ CISG-AC (n 118) para 3.1.

¹²⁵ Supreme Court 2 April 2009 <<http://cisgw3.law.pace.edu/cases/090402a3.html>> accessed 21 March 2016; Canton Appellate Court Vaud 11 April 2002 <<http://cisgw3.law.pace.edu/cases/020411s1.html>> accessed 21 March 2016; Tribunale di Forlì 11 December 2008 <<http://cisgw3.law.pace.edu/cases/081211i3.html>> accessed 21 March 2016; Decision 4505/2009 of the Multi-Member Court of First Instance of Athens 2009 <<http://cisgw3.law.pace.edu/cases/094505gr.html>> accessed 21 March 2016.

assumed if the corresponding intent of the parties is sufficiently clear.¹²⁶ The same outcome was reached in Switzerland in 2004, in a case where there was no exclusion of the convention because the parties had not stated their position.¹²⁷ The clear intent is also a requirement on the other side of the Atlantic. US courts appear to deny the possibility of implied opt-outs.¹²⁸ They construe Article 6 by comparison with Article 3 ULIS and therefore reject implicit exclusions as Article 6 has not retained the express wording of Article 3. In *Asante technologies v PMC Sierra*,¹²⁹ an American court applied the CISG because the clause ('the validity and performance of this order shall be governed by laws of the state shown on the buyer's address on this order'¹³⁰) was not considered clear enough to mean exclusion.

Apart from the express exclusion explained above, a clear intention can consist in choosing the law of a non-contracting state. As the CISG would not be part of the set of laws of this country, it is obvious that none of its provisions would be referred to. This is often accepted as an implied exclusion.

A further way to opt out impliedly would be to rely on usages which exclude the CISG. According to Article 9, if the parties decide that the usages of their particular trade will govern their contract or they know usages that are widely used in their trade and these usages consist in applying a certain set of rules that has no link with the CISG (English law for example), then the CISG will be excluded.

Yet courts are more often divided on what does and does not constitute an exclusion of the convention than they agree on when opt-outs are valid. A first glance at this problem was underlined earlier (with the *Asante* case). These divergences are important in practice because a choice-of-law clause held to be invalid could have dramatic consequences for the parties. Despite

¹²⁶ Supreme Court 22 October 2001 <<http://www.cisg.law.pace.edu/cases/011022a3.html>> accessed 21 March 2016.

¹²⁷ Appellate Court Jura (n 119).

¹²⁸ Franco Ferrari, 'Homeward Trend and Lex Forism in International Sales Law' [2009] IBLJ 337.

¹²⁹ Federal District Court [California] 27 July 2001 <<http://cisgw3.law.pace.edu/cases/010727u1.html>> accessed 21 March 2016.

¹³⁰ *ibid.*

this importance, this is the area that raises most uncertainties as courts disagree, and different reasoning can be found around the globe. This problem is evidenced in cases where the parties choose the law of a contracting state. For some judges, such a choice allows an inference of exclusion of the convention if the parties state that the chosen law will apply exclusively.¹³¹ However, in most cases, a reference to the law of a contracting state is not enough to opt out, even though the parties would not be aware of the existence of the convention, as the CISG is part of the law in such countries. The reason is simple: the standard of a clear intent to exclude the CISG is required because the convention is an opt-out system and not an opt-in one. Therefore, the question to be asked is whether the parties have sought to exclude the application of the convention and not whether they wanted to apply it to their contract.¹³² When the parties identify the law of a contracting state as the law governing their commercial relationship, it is not necessarily useless. Though the convention will apply, domestic law will still be relevant for matters not dealt with by the CISG. Thus, such a choice still makes sense. The diversity of cases offers numerous hypothesis of implied exclusion. For instance, an Australian court held that the convention was excluded where the contract contained the following clause: ‘Australian law applicable under exclusion of UNCITRAL law’.¹³³ The same conclusion is reached when the clause selects ‘the law of a contracting state insofar as it differs from the law of the national law of another Contracting State’.¹³⁴ In those decisions, the intent of the parties is clear enough to presume that they wished to rely on another law than the CISG. But the problem is that it is a matter of interpretation. Some courts may find that the intent is unequivocal while others would still apply the CISG. For example, it was held

¹³¹ Supreme Court 17 December 1996 (*Ceramique Culinaire v. Musgrave*) <<http://cisgw3.law.pace.edu/cases/961217f1.html>> accessed 21 March 2016; Bundesgerichtshof 11 May 2010 <<http://cisgw3.law.pace.edu/cases/100511g1.html>> accessed 21 March 2016.

¹³² Claude Witz, ‘L’Exclusion de la Convention des Nations Unies sur les Contrats de Vente Internationale de Marchandises par la Volonté des parties’ [1990] D 108.

¹³³ Federal Court of Australia 20 May 2009 (*Olivaylle Pty Ltd v. Flottweg GmbH & Co KGAA*) <<http://cisgw3.law.pace.edu/cases/090520a2.html>> accessed 21 March 2016.

¹³⁴ Oberlandesgericht [Appellate Court] Linz 23 January 2006 <<http://cisgw3.law.pace.edu/cases/060123a3.html>> accessed 21 March 2016; Austria (n 43).

that the incorporation of Incoterms does not amount to an exclusion of the CISG.¹³⁵ But another court found that it is an exclusion if provisions of Incoterms conflict with provisions of the convention.¹³⁶ In addition, the uncertainty is even higher when the parties are more precise. They can decide to choose the law of a contracting state but only a particular portion of it, such as a domestic statute regulating sales. The advisory council declared that ‘[a] reference to a Code containing the purely domestic sales law should be sufficient, provided the Code does not also enact the CISG’.¹³⁷ As a consequence, identifying a particular statute as the applicable law is a valid way to opt out. But it seems that the statute in question must be identified. It is not sufficient to refer to ‘domestic law’ in general according to a Belgian case,¹³⁸ though an Italian court has decided against this opinion and held that in such a case, the CISG was excluded.¹³⁹ One can see that implicit opt-outs may be understood differently depending on the jurisdiction.

Given the multiple ways to opt out, it is not surprising that traders are tempted to exclude the CISG. Moreover, the implied opt out favours the exclusion of the applicability of the CISG. Thus, the apparent simplicity of Article 6 conceals more problems. The CISG itself contains the very tools to lead to its destruction. Opting out is thus eased by the convention itself and this has for impact to penalise harmonisation of trade contract law.

¹³⁵ District Court Kortrijk 19 April 2001 (*CE v. CD and AR*) <<http://cisgw3.law.pace.edu/cases/010419b1.html>> accessed 21 March 2016.

¹³⁶ Cour suprême du canton [Appellate Court] Berne 19 May 2008 <<http://cisgw3.law.pace.edu/cases/080519s1.html>> accessed 21 March 2016.

¹³⁷ CISG-AC (n 116) para 4.5.

¹³⁸ Appellate Court Gent 20 October 2004 (*NV Van Heygen Staal v. GmbH Stahl- und Metalhandel Klockner*) <<http://cisgw3.law.pace.edu/cases/041020b1.html>> accessed 21 March 2016.

¹³⁹ Tribunale di Forlì 26 March 2009 <<http://www.globalsaleslaw.org/content/api/cisg/display.cfm?test=2336>> accessed 21 March 2016.

Then, another issue is the arguments used by the parties during litigation. Some litigants invoke only pure domestic provisions although the CISG would apply to their contracts. Several courts have held that it is not enough to opt out of the CISG.¹⁴⁰ Some courts went even further and adopted the same point of view even though the parties did not know that the CISG applied to their contract.¹⁴¹ But there remains a certain doubt on this issue as opposite decisions can be found. For example, the French Supreme Court held twice that the convention applied when parties based their arguments only on domestic law.¹⁴² This issue illustrates the complexity that can arise from a simple two-line article.

To conclude, as the name of the article¹⁴³ written by Drago and Zoccolillo suggests, parties should be explicit when it comes to the choice of law. The authors of this article declared that this will ‘ensure that the law intended by the parties is solidly established and not second guessed by courts’.¹⁴⁴

So, careful traders and conscientious lawyers should be sure that the mutual agreement to exclude the application of the CISG is expressly set out in the contract in order to avoid difficulties.

¹⁴⁰ District Court Saarbrücken 2 July 2002 <<http://cisgw3.law.pace.edu/cases/020702g1.html>> accessed 21 March 2016.

¹⁴¹ Appellate Court Zweibrücken 2 February 2004 <<http://cisgw3.law.pace.edu/cases/040202g1.html>> accessed 21 March 2016.

¹⁴² Supreme Court 25 October 2005 <<http://cisgw3.law.pace.edu/cases/051025f1.html>> accessed 21 March 2016; Supreme Court 26 June 2001 (*Anton Huber v. Polyspace*) <<http://cisgw3.law.pace.edu/cases/010626f1.html>> accessed 21 March 2016.

¹⁴³ Thomas J. Drago and Alan F. Zoccolillo, ‘Be Explicit: Drafting Choice of Law Clauses in International Sale of Goods Contracts’ (May 2002) 9 *The Metropolitan Corporate Counsel*.

¹⁴⁴ *ibid.*

If the possibility to opt out of the CISG is expressly contained in the convention, the ways to do that are not. The CISG offers no guidance as to the means to opt out. As Grave said in ‘Article 6 and issues of formation’:

“It was decided to address the issue of excluding [...] in a single sentence, leaving a simple and elegant statutory provision but one that does not clearly and unequivocally answer the question at hand.”¹⁴⁵

This statement underlines the fact that Article 6 is rather general. This may be explained by the desire to respect the party autonomy but also by the fact that the convention is a compromise between several legal systems, such as common law or civil law. Consequently, Article 6 is vague. It expresses the possibility to opt out but the ways to do so remain unexplained. Yet, that does not mean that it is complicated to opt out. If, as shown in the first part, opting out is so often applied, then it should not be difficult to do so as so many traders exclude the convention. A priori, it is relatively easy to opt out: one just has to say that the CISG is excluded and it is actually excluded. But one more time, this is not said in Article 6. So, the problem is not that the law is too complicated, because to be complicated it should at least provide some guidelines; but rather that the law is almost lacking. Nevertheless, if the wording of Article 6 is as it is, this is not without reasons. It may be seen as an attempt to give a choice to the parties. Opting out means to make the choice to withdraw from the effect of the provisions of the convention. This choice would not be complete without a free choice as to the ways to opt out. Moreover, it is impossible, for the sake of clarity, to list all the manners that exist to opt out. Giving the parties an open option to opt out as they wish is more consistent with the principle of party autonomy that emerges from the CISG. It is worth mentioning that this article is compliant with the principle of contractual freedom. Thus, it is consistent with the need for flexibility felt in international trade. However, more problems arise

¹⁴⁵ Jack Graves, ‘CISG Article 6 and Issues of Formation: the Problem of Circularity’ [2011] *Annals FacL.BelgradeIntlEd* 127.

from this broad article than solutions. As seen above, there are many uncertainties in the field of opt-outs.

To make the problem even thicker, if the law is not complicated, case law is. As there is no harmonisation of the decisions, traders and their lawyers are always in doubt as to the validity of the clause of choice-of-law. Each state (and even each court within a state sometimes) can have their own view on issues as there is no “international supreme court” to give uniform guidelines. The goal of harmonisation aimed at by the CISG seems to fade away.

Thus, small businesses (and even larger ones to some extent) which are not used to reading the law and do not have access to sharp and professional legal advice are left clueless as to how to opt out of the CISG. Many other problems are brought by opting out. They will be examined in the third part.

III. Difficulties in opting out: A plea for effective solutions

As seen earlier, there can be many reasons to exclude the CISG from international sale of goods contracts and many ways to do it. But there are also many difficulties brought by this choice. The consequences can be dramatic for parties and therefore opting out should be considered carefully. This part will study the problems raised by opting out, and then will focus on ways to improve Article 6 and defeat its complications.

The presence of Article 6 in the CISG has indisputable advantages. First of all, it respects freedom of contract. The parties are not compelled to be subjected to its rules. Following from that, the parties can choose a law they understand better or a law which is more favourable to them. Thus, opting out can decrease some costs. The costs of becoming familiar with the CISG may be an example: there will be no need to invest in knowledge if there is nothing new to learn. Opting out of the convention may also bring unification as to the rules applied to the contract. As the CISG does not cover every aspect of a contract, another law is required to fill in its gaps. If the parties choose to apply this law to the whole contract (and not a mix of national law and CISG), there will not be any inconsistencies in the resolution of the disputes. The rules will

come from the same source. In addition, opting out may give greater certainty about future conflicts. As the applicable law chosen by the parties is often a domestic law, the parties can benefit from a wide range of cases and interpretations of the provisions already well settled. They can have a better idea of how the dispute will end.

If some parties believe that opting out is the easiest way to avoid any trouble with the CISG, they may be surprised to see how many more problems they can attract by opting out. First, if the parties decide to opt out, they may have to agree on another applicable law. This may be complicated as hours of negotiation may be required to reach an agreement, which additionally means further costs. In most cases, each party will try to impose the law of his country. As a result, an agreement is often difficult to find. On the contrary, the CISG is ready to get picked up and offers a compromise that would decrease the time and costs linked to negotiation.¹⁴⁶ Besides, this agreement may consist in a compromise that does not fully satisfy both parties. To conclude with the issue of agreement, opting out may become problematic in case of a battle of the forms. This situation arises for instance when a seller sends an offer to sell goods; the offer including his own terms and the buyer accepts to buy the goods, but when sending his acceptance, he also addresses different contractual clauses. It might happen that a party does not read the other party's clauses and thus believes that the contract has been concluded on its terms. Yet, if a conflict arises later in their contractual relationship, there will be a problem: according to which forms should the dispute be dealt with? This is how a battle of the forms appears. For example, a party may choose the law of a particular contracting state and expressly excludes the CISG in his terms, but the other party, in his own terms, may make reference to the law of another contracting state. For the Advisory Council, it is not a case of opt out.¹⁴⁷ The Advisory Council calls upon the Draft Hague Principles on Choice of Law to bring light on the question. The solution would be that:

¹⁴⁶ Butler (n 100) ch 2.

¹⁴⁷ CISG-AC (n 116) 4.14.

“one should compare the battle of the forms rules under the CISG (either last shot or knock out) with the rule under [the law of the state chosen by the party who excluded the CISG] (excluding the CISG). If [the latter law] operates under a knock out rule, then both choices of law are knocked out, leaving no choice of law, and thus CISG applies.”¹⁴⁸

With this example, it is clear that playing with opt-out clauses may be a dangerous game. But this is not over. What if the opt-out clause is contained in general conditions which the other party has not read? Or what if the opt-out clause conflicts with standard terms? As Lavers pointed out, ‘[t]he choice of law that prevails will depend on which form prevails and, in turn, which form prevails may depend directly on whether you apply the rules of the CISG or the [domestic law]’.¹⁴⁹ Two problems emerge from that statement. First, it seems paradoxical to invoke the rules of a certain convention to enforce its exclusion. Article 19 would apply and may lead to the exclusion of the CISG. Secondly, it brings a question of circularity. To determine the applicable law, one needs to know which form has priority, but to know that, one must establish what is the applicable law. Even where the terms exchanged both express the intention of the parties to opt out of the CISG, it does not mean that the convention is excluded. The Advisory Council gives a relevant example:

“One party may indicate a choice of Danish law excluding the CISG, and the other might indicate Spanish law excluding the CISG. Neither Danish nor Spanish law will be applicable pursuant to the knock out method of dealing with conflicting standard terms expressed in CISG Advisory Council Opinion No 13, Rule 10.”¹⁵⁰

¹⁴⁸ *ibid.*

¹⁴⁹ Lavers (n 99) 12.

¹⁵⁰ CISG-AC (n 116) 4.14.

So, opting out may eventually have a reverse effect and lead to the application of the convention.

Then, if a choice of law clause is finally included in the contract, more problems may appear. The clause may be considered ineffective.¹⁵¹ As seen above, courts are divided as to how to construe opt-out clauses. As a consequence, even in cases where the parties have taken a common decision on the applicable law, it may not be enough to exclude the CISG.

Nevertheless, the main drawback of the popularity of opt-outs is that it makes the uniformity of cross-border business a utopia: often dreamt of but never reached. As opting out is so popular, there is no harmonisation of the laws of international trade transactions. Harmonisation is the main reason why the CISG is so important: it provides for a tool every trader, regardless of their nationality, can understand and use. The convention thus constitutes a common basis for international exchanges and should be used in order to facilitate these exchanges. However, this point of view does not seem to be shared by most traders.

The situation is indeed even less simple than it seems. Opting out too often creates a vicious cycle. If the culture of opting out expands, then there is no uniformity. If there is no uniformity, there is no certainty. Consequently, the parties prefer to opt out to avoid this uncertainty. And the cycle goes on and on again. Therefore, the familiarity with the rules of the convention develops and the CISG remains almost unknown. As Spagnolo said, 'opt-out contributes to unfamiliarity through a lack of exposure to CISG dispute work'.¹⁵² Besides, Article 6 encourages forum shopping¹⁵³ whereas a unified contract law would avoid this problem. But this problem cannot be erased from the CISG otherwise it would deny the parties the right to freely choose how to draft their contract.

¹⁵¹ Butler (n 100) ch 2.

¹⁵² Lisa Spagnolo, 'The Last Outpost: Automatic CISG Opt Outs, Misapplications and the Costs of Ignoring the Vienna Sales Convention For Australian Lawyers' (2009) 10 *MelbJIntLaw* 141 pt IV.

¹⁵³ Ferrari, 'Forum Shopping' (n 89).

With so many drawbacks and a law that is so vague that it makes it complicated to opt out, it is high time to find effective solutions. Potential solutions can be suggested to improve Article 6 and ease opt-outs.

Firstly, it is worth underlining that case law gives increasing guidelines for opting out. The first thing to do to have more details about Article 6 and the way opt out is applied in practice is to read decisions. As the CISG has been used for almost 30 years, judges have come across cases where the CISG had to be implemented. Therefore, they gave practical examples which can be read as a guide. They explain what to do and what not to do to effectively opt out. Nevertheless, judgments differ from one country to another. If they can give indications, they are not to be entirely trusted. A French seller could be aware of a decision from an American court about a certain interpretation of an opt-out clause but could not completely rely on it as French judges may have a different opinion. Ferrari stressed the fact that ‘courts seem to resort to nationalistic interpretations’.¹⁵⁴ Thus the many decisions dealing with the CISG and easily available in databases should only be regarded as guidelines and not as pure laws. Nonetheless, case law could become a real cornerstone of the application of the CISG. Ziegel suggested establishing an international tribunal with authority for interpretation.¹⁵⁵ Functioning on the model of the CJEU, judges from all states could refer to that tribunal when they are unsure of how to tackle the issue at stake. The decisions of the tribunal would be binding on all contracting states, creating a unified sphere of common interpretations. Bonell had a similar idea. He proposed to create a board with representatives of each contracting state.¹⁵⁶ Based on analysis of these representatives, the board would draw a comparative study and the UN could revise the convention in accordance with this study.

¹⁵⁴ Ferrari, ‘Homeward trend’ (n 128) 334.

¹⁵⁵ Jacob S. Ziegel, ‘The Future of the International Sales Convention from the Common Law Perspective’ (November 2000) 6 NZBLQ 345.

¹⁵⁶ John E. Murray, ‘The Neglect of CISG: a Workable Solution’ (1998) 17 JLC 365.

But what if the solution to the vagueness of Article 6 was already in the convention? Perhaps one has to read between the lines to explain how opt-outs work. Article 6 has to be understood in the context of the CISG, as a part of the whole convention. Therefore, its interpretation should be examined under the lights of the spirit that emerges from the entire text. It should be guided by the general principles of the convention. For instance, when two parties have included an unclear opt-out clause in their contract, the judge in charge of the matter could use the principle of party autonomy as a justification for his decision. As party autonomy is one of the general principles, it could be used to support the choice of the parties to contract out even though their intention was not very clear. So to some extent general principles might be helpful. General principles have been discussed at length and some authors have given them some weight to interpret the CISG. Bailey developed an interesting reasoning. He said that one should start from the language of the CISG, then go to case law, and finally move to unstated principles.¹⁵⁷ Thus good faith or fair trading could help to interpret Article 6 in a given context. Bailey also stated that to understand the CISG, the secretariat commentary should be adopted as official commentary.¹⁵⁸ The commentary could constitute the guidelines the business world has been looking for.

Then, if it is not enough to look at decisions and principles, the CISG could be changed into an opt-in instrument. This would have the advantage of facilitating the interpretation of intention of parties in cases where the choice of law clause is unclear. However, it would not favour harmonisation of transactions rules across the world. Yet, as the CISG is often excluded, one may think that this handicap is of little weight and that it would be more useful for traders to have the possibility to opt in instead of opt out.

¹⁵⁷ James E. Bailey, 'Facing the Truth: Seeing the Convention on Contracts for the International Sale of Goods as an Obstacle to a Uniform Law of International Sales' (1999) 32 *CornellIntlLJ* 297.

¹⁵⁸ *ibid* 299.

In conclusion, whether the law ought to be changed or not is a matter of policy. If one supports the CISG, then it is obvious that the convention must not become an opt-in legislation. If one disagrees with this point of view, the convention should be turned into an opt-in instrument. However, in both cases, it is important to at least create effective guidelines on Article 6 for courts of the world.

Nonetheless, it seems impossible to suppress all the areas of doubt surrounding Article 6. Thus, the ultimate solution would not be one to facilitate opt-outs but to reduce their popularity. By promoting the CISG, traders and legal professionals would become more familiar with the convention and would be less reluctant to apply it. Thus, they may give up implementing opt-out clauses. Education plays an important role here. The more traders know about the CISG, the more they are able to understand when it is better for them to apply it or not. Explanations of the mechanisms of the convention and precisions on how to opt out in an efficient way could be provided by chambers of commerce, for example, as long as they give a uniform interpretation. In addition, traders should not be the only ones to receive an education about the CISG. Lawyers should also be more aware of the convention. As they advise clients, they have a duty to know the relevant legislation in order to develop a strategy that protects the best interests of clients, and accordingly the CISG is amongst legislation they should be aware of. But this is the theory. In practice, many lawyers know little about the convention. They may know that it exists but have no idea as to how it works.¹⁵⁹ Schwenzer and Kee underlined that ‘there are many degrees of familiarity, and it would appear unwise to equate familiarity with a genuine understanding of how the CISG operates and can operate in international trade’.¹⁶⁰ Several solutions could increase the awareness of lawyers. First, the CISG could be taught more in law schools. In a survey realised in the USA, Fitzgerald showed that most respondents agreed that the CISG should be taught more.¹⁶¹ The same survey brings hope because it estimates that 78%

¹⁵⁹ Schwenzer and Kee (n 90) 433.

¹⁶⁰ *ibid* 438.

¹⁶¹ Fitzgerald (n 90) 10.

of professors address the CISG in contract courses and that 100% of them teach it in sales law courses.¹⁶² Thus, there seems to be a good cover of the CISG in law schools. Dodge makes the reverse observation and claims that the CISG is neglected by casebooks.¹⁶³ This indicates there may be some greater effort needed. Secondly, the CISG could be made mandatory at the bar exam. Future lawyers would thus be aware of the mechanisms of the CISG and would be able to implement it once they become practitioners. However, for the picture to be complete, judges have also a few lessons to take. As judges may not deal with the convention on a daily basis, it is arguable that when a case comes into their hands they may not be as confident to apply the convention as they are with domestic law. So, they can make mistakes in interpreting the CISG. In addition, judges are always lacking time to analyse cases.¹⁶⁴ Sometimes they cannot afford to spend hours reading cases in order to interpret a single clause of a contract, and they have even less time to read CISG cases from outside their jurisdiction.¹⁶⁵ The solution would thus be to make judges more familiar with the CISG from the very beginning, so that they would not have to spend that extra time in research. They could be more efficient and find a fair answer faster. Consequently, education is a key element to promote the CISG but also to prevent difficulties in opting out.

Another way to limit the problems linked to opt-outs would be to improve the CISG so that traders would want to apply it rather than exclude it. In order to do so, Ziegel suggested making amendments to the original text of the convention to enable its adjustments to the current need of traders.¹⁶⁶ If the CISG stays up to date with the latest practices, then traders would see the convention could benefit them. Thus, they would save themselves the trouble of opting out by having their transactions governed by the convention. This solution does not provide an answer to the difficulties brought by Article 6, but may be able to reduce the number of cases where these problems arise.

¹⁶² *ibid.*

¹⁶³ William S. Dodge, 'Teaching the CISG in Contracts' (2000) 50 *JLegalEduc* 76.

¹⁶⁴ Schwenzer and Kee (n 90) 445.

¹⁶⁵ *ibid.*

¹⁶⁶ Ziegel (n 155) 445.

Thus, the solution which consists in improving the CISG to decrease the proportion of opt-outs may be more in line with the goal of the CISG. To harmonise international trade transactions, it seems more logical to make the CISG more attractive and easier to understand. Consequently, it would finally be up to the parties to decide whether the CISG suits their interests or not. Article 6 will always enable them to opt out. And, through a better education, they will know how to do that.

To conclude, one cannot say that the law is too complex. It is true that traders do not like complicated legal vocabulary. They rather prefer straightforward provisions so they can secure their transactions. They also like having the choice: the choice to choose the applicable law, the choice to include their own conditions, the choice to change ineffective clauses, and so on. Article 6 gives them this freedom but does not give them the clarity they long for. Article 6 enables them to opt out of the CISG but does not say how to do so. What is unique with Article 6 is that many surveys pointed out that opting out is rather popular in practice despite the lack of clarity of the convention. It seems that traders prefer to challenge the obscurity of the law than putting their contract under the lights of the convention. Thus, several methods to opt out were given birth. From express choice of law clauses to implied choices, the imagination of traders (and their lawyers) has been prolific. However, this imagination has also driven them to unsettled paths. Through unclear clauses, traders expose themselves to a potential invalidity of opt-out clauses. And this is because the law is vague. Article 6 is too concise to explain what is considered a good opt-out and what is not. Therefore, the law is not too complicated but rather almost empty. It should not remain so. Because opting out may bring problems, a better understanding of Article 6 is required. Several solutions may be envisaged but will only be achieved through an international cooperation. This would be beneficial for the different actors of the international business world but would also have a greater impact: this could develop cross-border exchanges thanks to a more uniform legal context and could stimulate the worldwide economy. However, the best solution would be to be patient. As more and more cases apply the CISG and the current trend reflects

a decrease in opt-outs, a ‘change has begun’.¹⁶⁷ One can hope this will lead to a clearer understanding of Article 6 without requiring a complex amendment of the text.

¹⁶⁷ Harry M. Flechtner, ‘Changing the Opt-out Tradition in the United States’ (Congress to celebrate the fortieth annual session of UNCITRAL, Vienna 9-12 July 2007).

Class action (not US-style): enhancing access to justice

Shinthean Ng¹⁶⁸

'A class action by consumers produces several salutary by-products, including a therapeutic effect upon those sellers who indulge in fraudulent practices, aid to legitimate business enterprises by curtailing illegitimate competition and avoidance to the judicial process of the burden of multiple litigation involving identical claims.'

- *Vasquez v Superior Court*¹⁶⁹

In 2008, Mulheron produced a research paper¹⁷⁰ which was then submitted to the Civil Justice Council (CJC). It aimed to give evidence for the need for a new class action regime. The CJC later produced a report¹⁷¹ which endorsed her proposals on having an opt-out class action regime. However, the Ministry of Justice (MoJ) rejected the recommendation.¹⁷² More recently, the Consumer Rights Act 2015 created an opt-out class action¹⁷³ in the area of competition law. This opens up the debate for this essay to discuss whether some type of class action procedure must be present in order to protect consumer rights adequately.

¹⁶⁸ LL.B. Candidate, The University of Manchester, School of Law.

¹⁶⁹ (1971) 4 C3d 800.

¹⁷⁰ 'Reform of Collective Redress in England and Wales: A Perspective of Need' (2016) <www.judiciary.gov.uk/wp-content/uploads/JCO/Documents/CJC/Publications/Other+papers/reform-of-collective-redress.pdf> accessed 13 May 2016.

¹⁷¹ "Improving Access to Justice through Collective Actions' Developing a More Efficient and Effective Procedure for Collective Actions' (2016) <www.ucl.ac.uk/laws/judicial-institute/files/Improving_Access_to_Justice_through_Collective_Actions_-final_report.pdf> accessed 13 May 2016.

¹⁷² 'Reforming English Civil Procedure: Class Actions and Third Party Funding' (2016) <www.law.qmul.ac.uk/research/impact/class_actions/index.html> accessed 13 May 2016.

¹⁷³ Consumer Rights Act 2015, s 81.

I. Introduction

At the outset, a case for collective redress will be established, reflecting the current problems in the collective redress options available to consumers in light of the opt-in system and funding issues. Subsequently, it will be further proved how a class action vehicle can solve these problems. Ultimately, this essay will endorse the view that, consumers are in need of some type (not US-style) of class action court procedure to obtain adequate protection for their rights.

II. Making a case for collective redress

A survey conducted by Eurobarometer in 2004 revealed that, in fifteen EU Member States included in its study, 67 per cent of consumers would be more willing to defend their rights in the court if they were able to 'join with other consumers who were complaining about the same thing'.¹⁷⁴ The case *Meux v Maltby*¹⁷⁵ also expressed the 'great inconvenience' of bringing all the parties before the court. Especially in product liability cases involving design defects, a wide range of consumers may be affected by a single act on the part of a manufacturer.¹⁷⁶ Additionally, in consumer credit cases, contract forms are standardised in mass transactions and creditors often deal with consumer debtors in an identical fashion.¹⁷⁷ On this account, a point is made to adopt some sort of collective redress mechanism. Since court action is sometimes the only way for consumers to seek justice,¹⁷⁸ it will be argued that adopting an opt-out class action would be in the best interest of consumers.

¹⁷⁴ Sutatip Yuthayotin, Access to Justice in Transnational B2C E-Commerce 189.

¹⁷⁵ [1812] 2 Swans 277 at 281.

¹⁷⁶ D W Oughton and John Lowry, Textbook on Consumer Law (Blackstone 2000) 86.

¹⁷⁷ Mark Mildred, 'The Reform of Class and Representative Actions in European Legal Systems: A New Framework for Collective Redress in Europe by Christopher Hodges' (2009) 72 Modern Law Review 287.

¹⁷⁸ 'Representative Actions in Consumer Protection Legislation Consultation' (2016) <<http://webarchive.nationalarchives.gov.uk/20070603164510/http://www.dti.gov.uk/files/file31886.pdf>> accessed 13 May 2016.

III. Making a case for class actions

I) Difficulties of Representative Action

One possible way to encourage private actions while avoiding some excesses of the US-style class action is through the representative action that falls under Rule 19.6 of the Civil Procedure Rules.¹⁷⁹ However, Neil Andrews observed that the representative action remains ‘distinctly marginal’ in England due to the ‘loser pays’ rule.¹⁸⁰ Since represented parties are technically not parties to the representative action, the representative that brings the action would be liable for all of the costs incurred if the case is lost.

Other serious disadvantages of this procedure are the requirements of ‘more than one person’ and a ‘same interest’ in claim.¹⁸¹ In this regard, the Scottish Law Commissioners thought that it was unclear as to what constitutes ‘numerous persons’ having the ‘same interest’.¹⁸² Moreover, the definition of ‘same interest’ has been narrowly limited by the interpretation of the English Court of Appeal in *Markt & Co Ltd v Knight Steamship Co Ltd*.¹⁸³ It was held that separate contracts between the defendant and each of the class members, even if the contracts were in *identical* terms, would not satisfy the requirement. In another case, *Emerald Supplies v British Airways*,¹⁸⁴ the Court of Appeal rejected the representative action as being improperly constituted. The ruling left the commentators wondering how consumers will gain access to justice.¹⁸⁵

¹⁷⁹ ‘Part 19 - Parties and Group Litigation - Civil Procedure Rules’ (justice.gov.uk, 2016) <www.justice.gov.uk/courts/procedure-rules/civil/rules/part19#19.6> accessed 13 May 2016.

¹⁸⁰ Neil H Andrews, ‘Multi-Party Actions and Complex Litigation in England’ SSRN Electronic Journal 259.

¹⁸¹ Civil Procedure Rules, Rule 19.0.10-19.15.

¹⁸² ‘Class Actions in the Scottish Courts: A New Way for Consumers to Obtain Redress?’ (2016) <<http://webarchive.nationalarchives.gov.uk/20090724135150/http://scotcons.demonweb.co.uk/publications/reports/documents/ClassActionsintheScottishCourts.Anewwaytoobtainredress.May1982.pdf>> accessed 13 May 2016.

¹⁸³ [1910] 2 KB 1021.

¹⁸⁴ [2015] EWCA Civ 1024.

¹⁸⁵ Katy Dowell, ‘Class Dismissed but Hausfeld to Appeal Again’ (2010) 24 The Lawyer 6.

II) Difficulties of Group Litigation Order (GLO)

Given the shortcomings of the representative actions, amendment to the Civil Procedure Rules was made in 2000 to allow the courts to issue GLOs.¹⁸⁶ Due to the nature of this procedure, it has been used primarily for high value disputes and it is not appropriate for cases where there are many small economic losses.¹⁸⁷ The opt-in requirement attached to GLOs means that they are unlikely to be used for cases involving mass consumer losses where individuals face significant barriers to 'opting-in'. This perhaps explains, for example, why there is only one case to date which concerns a claim made by the Consumers' Association in relation to price fixing in replica football kit.¹⁸⁸

III) Introducing a Class Action

Mulheron stated in her report that there is a gap in the collective redress mechanisms available in England, and a more effective method of collective redress is urgently required to address it.¹⁸⁹ She used the case of *Independiente Ltd v Music Trading On-line (HK) Ltd*¹⁹⁰ to illustrate that representative action in Civil Procedure Rules is developing in order that the introduction of a class action may not be too dramatic from the UK's perspective.¹⁹¹ Thus, the possible solution so far seems to be to create an opt-out class action to overcome the deficiencies of the current collective redress schema.

¹⁸⁶ Edward F Sherman, Group Litigation under Foreign Legal Systems (2002) 52 DePaul Law Review 2, 423-24.

¹⁸⁷ *ibid*, 424.

¹⁸⁸ 078/7/8/07 *The Consumers Association v JJB Sports PLC*.

¹⁸⁹ (n 170) 157-61.

¹⁹⁰ [2007] FSR 21 (Ch).

¹⁹¹ 'Evaluation of the Effectiveness and Efficiency of Collective Redress Mechanisms in the European Union' (2016) <http://ec.europa.eu/consumers/archive/redress_cons/uk-country-report-final.pdf> accessed 13 May 2016.

Despite the argument that class action provides greater access to the courts,¹⁹² it is frequently criticized for creating an unsavoury litigation culture, as is the case in the US jurisdiction.¹⁹³ Abuse of the court procedure can prevent substantive consumer law from developing,¹⁹⁴ thereby backfiring against the purpose of enhancing consumers' access to justice. Additionally, it can be argued that class action does not itself fit well within the English context, as per the Irish context, due to different legal cultures.¹⁹⁵ Apart from that, the Small Claims Procedure has allowed smaller claims by consumers to be litigated at lower costs.¹⁹⁶

Mulheron, the forefront proponent for class action, has been challenging criticisms all the while. She contended that the reasons given (judicially and academically) to reject a class action lack 'cogency and substance' in evaluating suitable models.¹⁹⁷ Firstly, it seems that this jurisdiction has been overly influenced by the extremism of the US-style actions. By way of contrast, there are few successful examples in other jurisdictions which will be discussed below. Furthermore, research on small claims in England proved that many potential litigants are reluctant to embark even on the simplified procedures offered by the small claims courts.¹⁹⁸ As a matter of fact, the allegation that there is a lot of unnecessary litigation is inconsistent. There is little evidence from the US that *certified* class actions were mostly frivolous. Recent evidence from and regarding consumer class actions against insurance companies suggest that the judiciary effectively screens out meritorious suits.¹⁹⁹ Even the Scottish Law Commissioners rejected the argument that a class action will necessarily encourage litigation culture.²⁰⁰

¹⁹² Rachael P Mulheron, *The Class Action in Common Law Legal Systems* (Hart 2004) 9-11.

¹⁹³ (n 182) 23.

¹⁹⁴ Mulheron (n 193) 31.

¹⁹⁵ (n 182) 27.

¹⁹⁶ 'A Class of Their Own: Why Scotland Needs a Class Procedure?' (2016) <<http://webarchive.nationalarchives.gov.uk/20090724135150/http://scotcons.demonweb.co.uk/publications/reports/reports03/rp09aclass.pdf>> accessed 13 May 2016, 7.

¹⁹⁷ Mulheron, Rachael. 'Some Difficulties with Group Litigation Orders - and Why a Class Action is Superior' (2005) *Civil Justice Quarterly* 40.

¹⁹⁸ Mulheron, Rachael (n 197) 7.

¹⁹⁹ Geraint G Howells and others, *Handbook of Research on International Consumer Law* (Edward Elgar 2010) 533-34.

²⁰⁰ Mulheron, Rachael (n 197).

Furthermore, importing class action regime does not mean that we have to implement the US-style litigation.²⁰¹ Sir Anthony Clarke once said, 'It is clearly not desirable to import a US-style class action system, nor would it be practical to do so.'²⁰² City law firm, Clifford Chance, expressed in its briefing note that:

*'Although [the] features [of collective actions in the UK] have probably curbed any US-style class action explosion, the UK can expect to see an increase in collective actions.'*²⁰³

In order to see an increase in collective actions by consumers, some type of class action has to exist. Successful class action regimes require modifying the class action regime itself in order to suit the legal culture, history, constitution, court system and rules governing the legal profession in the jurisdiction.²⁰⁴ With regard to this, it is submitted that an opt-out system is mandatory in order for access to justice to function.

i) Opt-Out System

One practical problem of an opt-in system under the current GLO is shown by *Taylor v Nugent Care Society*²⁰⁵ (although it is not a consumer case). The case demonstrates that a victim who missed the deadline for joining GLO would have problems bringing a claim to the court and that this procedure is wasteful of the litigant's resources.²⁰⁶ Conversely, in an opt-out regime, there is no pressure to ensure that all claimants issue individual proceedings in the very early stages so as to join the group register. In consequence,

²⁰¹ Gidi, Antonio, 'Class Actions in Brazil: A Model for Civil Law Countries' (2003) 2 The American Journal of Comparative Law 51, 311-408.

²⁰² 'CJC Response - February 2008' (2016) <www.judiciary.gov.uk/related-offices-and-bodies/advisory-bodies/cjc/collective-action-reform/> accessed 13 May 2016.

²⁰³ 'Collective Actions in Europe' (2016) <www.cliffordchance.com/content/dam/cliffordchance/PDF/collective_actions_europe_2010.pdf> accessed 13 May 2016.

²⁰⁴ For example, Brazilian Consumer Code introduced a US type form of class action in 1990, with interesting and original procedural adaptations.

²⁰⁵ [2004] 1 WLR 1129.

²⁰⁶ Stephen C Yeazell, 'Group Litigation and Social Context: Toward a History of the Class Action' (1977) 77 Columbia Law Review 40.

participation rates are much higher in opt-out regimes (more than 87 per cent).²⁰⁷ An opt-out system is the norm in the US; Spain²⁰⁸ and Portugal²⁰⁹ already have opt-out systems, and even Hong Kong is considering an opt-out class action system for consumer cases.²¹⁰ Hence, if some type of class action is to be introduced, it should be based on an opt-out system to increase access to justice.²¹¹

It is also worth considering that some countries adopted a hybrid approach. This can be a measure to mitigate the possible consequences of frivolous, unjustified and disproportionate claims. For example, Norway's Dispute Act 2005 requires that class members opt-in under most circumstances, but allows class actions involving small and largely undifferentiated claims to proceed on an opt-out basis.²¹² Denmark allows a publicly appointed Ombudsman to bring opt-out class actions on behalf of consumers, but only allows private individuals to represent classes whose members have opted in.²¹³

As far as a class action is concerned, it is contended that, to enhance the protection for consumers, an opt-out procedure should be introduced. As mentioned above, this does not mean that the US system should be transplanted into our system. In countries like Portugal, the consumer organisation, DECO, thought that its class action regime had worked well.²¹⁴

²⁰⁷ (n 170) 7-8.

²⁰⁸ Law of Civil Judgment 1/2000.

²⁰⁹ Right of Proceeding, Participation and Popular Action No. 83/95.

²¹⁰ 'Class Actions - From North America to Europe and Beyond? Freshfields' (2016) <www.freshfields.com/foresight/article-03-html> accessed 13 May 2016.

²¹¹ Harry Woolf, Access to Justice: Interim Report to the Lord Chancellor on the Civil Justice System in England and Wales (Woolf Inquiry Team, 1995) 130.

²¹² 'Class Actions, Group Litigation & Other Forms of Collective Litigation in the Norwegian Courts' (2016) <http://globalclassactions.stanford.edu/sites/default/files/documents/Norway_National_Report.pdf> accessed 13 May 2016.

²¹³ Christopher J S Hodges, *The Reform of Class and Representative Actions in European Legal Systems* (Hart Publishing 2008) 34.

²¹⁴ (n 171).

ii) Funding

One of the major drawbacks under the existing procedures of representative action and GLO is the ‘loser pays’²¹⁵ rule. It is said that the rule can help to filter out unmeritorious claims.²¹⁶ However, the force of argument in favour of public funding is extremely strong, since consumer actions are brought on behalf of a large segment of the general public.²¹⁷ Some of the funding problems associated with the lack of contingent fee could be alleviated if the ‘loser pays’ rule is reformed to make the consequences less draconian.

Quebec has come close to this approach by limiting the class members’ liability to nominal costs; in Ontario, the Solicitors’ Act 2002 provided for regulated contingency fees in group actions and other complex cases (when no other methods of funding are available).²¹⁸ In this connection, the CJC suggested allowing *regulated* contingency fees under a class action as a last resort. Alternatively, there is third party funding which has become a feature of litigation in Australia.²¹⁹ In consideration of this method, the English Court of Appeal in *Arkin v Borchard Lines Ltd*²²⁰ held that:

‘We consider that a professional funder, who finances part of a claimant’s costs of litigation, should be potentially liable for the costs of the opposing party to the extent of the funding provided. The effect of this will, of course, be that...the net recovery of a successful claimant will be diminished. While this is unfortunate, it seems to us that it is a cost that the impecunious claimant can reasonably be expected to bear.’

²¹⁵ Fairgrieve D and Howells G, ‘Collective Redress Procedures’ (2009) 58 International and Comparative Law Quarterly 379, 395.

²¹⁶ Samuel Issacharoff and Geoffrey P Miller, ‘Will Aggregate Litigation Come to Europe?’ SSRN Electronic Journal 201.

²¹⁷ *ibid.*

²¹⁸ “‘Improved Access to Justice - Funding Options & Proportionate Costs’” (2016) <www.judiciary.gov.uk/wp-content/uploads/JCO/Documents/CJC/Publications/CJC+papers/CJC+Improved+access+to+Justice+-+Funding+options+and+proportionate+costs.pdf> accessed 13 May 2016, 39.

²¹⁹ *ibid.*, 41.

²²⁰ [2005] EWCA Civ 655.

In other words, in a case where the class action is funded by a third party, even though the damages received by the consumers would be lesser, it is nevertheless a means to facilitate access to justice if no other options are available.

It has been suggested that under the new class action regime, an opt-out system is mandatory, although it can vary flexibly according to different instances as exemplified by the Norway and Denmark models. On the issue of funding, it is recommended that the 'loser pays' rule should be substituted by regulated contingency fees or third party funding to encourage consumers to participate in class action.

IV. Conclusion

As demonstrated, representative action procedures and GLOs are of limited utility in ensuring adequate protection of consumer rights. On the positive side, class action is not the perfect procedure but it is the best alternative. The real message is to restructure the class action concept to reflect the weaknesses of the US-style class action²²¹ instead of rejecting the procedure itself as a means of enhancing consumers' access to justice. Thus, this essay has considered a regulated opt-out system, rules on contingency fees, and the third-party funding method.

The UK Consumer Rights Act 2015, which took the first step to introduce an opt-out class action procedure, has raised huge public concerns.²²² It is true that the fear of frivolous and vexatious litigation will continue. One way or another, class action procedure is here to stay and to protect consumer rights more adequately, as there is consensus that it will soon be the norm in Europe.²²³

²²¹ Cappalli, Richard B and Claudio Consolo. 'Class Actions for Continental Europe - A Preliminary Inquiry' 6 (1992) Temp. Int'l & Comp LJ 217.

²²² 'US-Style Class Actions Introduced in UK' (2016) <www.bbc.co.uk/news/uk-34402483> accessed 13 May 2016.

²²³ Samuel Issacharoff and Geoffrey P Miller (n 216) 287.

Dworkin's Hercules as a model for judges

Arvinth Rai²²⁴

In the 1975 paper *Hard Cases*, Ronald Dworkin first introduced Hercules, his well-known model of the ideal judge – a superhuman, omniscient judge of infinite competence, intelligence and resourcefulness.²²⁵ Hercules' profile is fleshed out in the 1986 book *Law's Empire*, where he develops into the protagonist of Dworkin's theory.²²⁶ This paper brings three "problems" to the fore, which together illustrate why Hercules does not provide a good model for real judges to emulate. The first is the problem of individuation of laws, which impairs Hercules' ability to construct a scheme of Dworkinian principles. The second is the problem of indistinguishability of criteria and the third is that of incommensurability of criteria, which both undermine Hercules' method of adjudication.

I. Introduction

Hercules, Ronald Dworkin's model of the ideal judge, first emerged in *Hard Cases* some 30 years ago to a mixed literary reception. Some described him as the alter ego of Ronald Dworkin himself,²²⁷ others likened him instead to Noam Chomsky's "ideal speaker-hearer",²²⁸ and yet others painted him as a "loner", who "converses with no one" and whose "narrative constructions are monologues".²²⁹ This paper contributes to the literature by placing Hercules under empirical lenses. The fundamental question we are concerned with relates to the usefulness of Hercules as a model for real judges to emulate. In answering this question, this paper will first briefly explain Dworkin's theory of law – particularly, his theory of adjudication – and his use of Hercules. Thereafter, some

²²⁴ LL.B. Candidate, The University of Manchester, School of Law.

²²⁵ Ronald Dworkin, 'Hard Cases' (1975) 88 Harvard Law Review 1057, 1083.

²²⁶ Ronald Dworkin, *Law's Empire* (Fontana 1986).

²²⁷ Allan C Hutchinson, *Evolution and the Common Law* (Cambridge University Press 2005) 82

²²⁸ Christopher Hutton, *Language, Meaning and the Law* (Edinburgh University Press 2009) 151

²²⁹ Frank I Michelman, 'Foreword: Traces of Self-Government' (1986) 100 Harvard Law Review 4, 76.

of the most important arguments against Hercules will be substantiated. Finally, Dworkin's defence of Hercules will be assessed.

II. The Dworkinian Enterprise

Dworkin's theory is premised on what he perceives as inadequacies in the theory of adjudication propounded by Hartian positivism.²³⁰ For Dworkin, Hart's characterisation of the law as a system of rules fails to account for legal principles. Hart's explanation of how judges decide hard cases – that is, they use their discretion to make new law – disregards the numerous cases wherein judges speak of their being bound by law even though no rules apply.²³¹

Dworkin thinks that judges rely and should rely on principles in hard cases.²³² His conception of law as integrity offers judges a blueprint for adjudication.²³³ It directs judges to decide cases by engaging in a constructive interpretation of legal practice and integrate their decisions with the existing body of law.²³⁴

What does Dworkin mean by this? Dworkin recognises the complexity of his theory and creates Hercules to articulate it. Hercules' preliminary task is interpretive. He must produce an entire theory of law by constructing a scheme of principles that justify all the rules of his jurisdiction. Against this background, Hercules, in *Law's Empire*, decides three types of cases: common law, statutory and constitutional cases. Our analysis of Hercules will focus on his interpretive task and his method of deciding common law cases. These two aspects of Dworkin's use of Hercules are most instructive in explaining how Hercules provides a model for judges. By constructing a scheme of principles, Hercules shows real judges the

²³⁰ Ronald Dworkin, *Taking Rights Seriously* (Duckworth 1977) 81; Michael Freeman, *Lloyd's Introduction to Jurisprudence* (9th edn, Sweet & Maxwell 2014) 313.

²³¹ See, for example, *Henningsen v Bloomfield Motors Inc.*, 161 A2d 69 (NJ 1960) discussed in Dworkin (n 230) 23-26 and Scott J Shapiro, 'The "Hart-Dworkin" Debate: A Short Guide for the Perplexed' (2007) University of Michigan Public Law Working Paper 77 <<http://ssrn.com/abstract=968657>> accessed 28 March 2015.

²³² Ronald Dworkin, *A Matter of Principle* (OUP 1985) 11.

²³³ cf integrity as a legislative principle: Dworkin (n 226) 176; Freeman (n 230) 599-600.

²³⁴ Dworkin (n 226) 226; Freeman (n 230) 599-600; Stephen Guest, *Ronald Dworkin* (Edinburgh 1992) 46.

form of decision-making they presuppose when deciding cases.²³⁵ He does fully and expressly what real judges do – that is, rely on principles – in a piecemeal and less conscious way. Meanwhile, in exhibiting how he decides cases, Hercules is effectively illustrating Dworkin’s theory of adjudication. His methods of deciding statutory and constitutional cases generally mirror his method of deciding common law cases.²³⁶

(i) Hercules’ Interpretive Task

Hercules interpretive task requires him to not only uncover principles but their associated sub-principles.²³⁷ Take, for example, the defence of fair comment in the law of defamation. Hercules may justify fair comment on the basis that people are entitled to exercise their freedom of speech to make comments on matters of public interest. There are further principles which will underpin what constitutes a “comment”. For example, the principle that there is a moral difference between *informing* and *critiquing* may justify why the fair comment defence does not extend to comments presented as statements of facts. Note that we are free to uncover principles that we consider better constructive interpretations of the fair comment defence. For instance, we may prefer instead the principle that people have no moral right to compensation from those who state views that a reasonable person may hold.

²³⁵ Dworkin (n 226) 265; Nigel E Simmonds, *Central Issues in Jurisprudence* (4th edn, Sweet & Maxwell 2013) 207.

²³⁶ Particularly statutory cases: Dworkin (n 226) 313, 337, 397.

²³⁷ Guest (n 234) 48.

(ii) Hercules' Method of Adjudication

How does Hercules then decide common law cases? To explain, we must first consider the process of interpretation.²³⁸ At the interpretive stage, Hercules sets out candidates for the best interpretation of the precedent cases.²³⁹ As a judge committed to integrity, he must think of himself as an author in the “chain of common law” and seek an interpretation that *fits* with the constraints existing in the chapters of the law written before him.²⁴⁰ In pursuit of fit, Hercules will turn to other areas of law or, if necessary, his entire theory of law.²⁴¹ Thus, at the interpretive stage, fit is essentially a threshold requirement which any eligible interpretation must meet.²⁴²

Hard cases arise when two or more interpretations satisfy Hercules' convictions about fit.²⁴³ Hercules must then decide which interpretation *appeals* most, taking the narrative forward according to justice, fairness and procedural due process.²⁴⁴ He must also, at this post-interpretive stage, bear in mind the criterion of fit. At this stage, fit plays a more evaluative role of determining what the best interpretation is.²⁴⁵

III. Attacking Hercules

This paper's critique of Hercules is twofold. It will be argued that Hercules does not provide a good model for real judges because his interpretive task does not reflect what real judges do in hard cases, and real judges cannot and should not emulate his method of adjudication.

²³⁸ *ibid* 65-68; Andrei Marmor, *Interpretation and Legal Theory* (2nd edn, Hart 2005) 54.

²³⁹ There is also a pre-interpretive stage, but it is not relevant for our purposes.

²⁴⁰ Dworkin (n 226) 238-239.

²⁴¹ Dworkin (n 230) 117; JW Harris, 'Unger's Critique of Formalism in Legal Reasoning: Hero, Hercules, and Humdrum' (1989) 52 MLR 42, 51.

²⁴² Dworkin (n 226) 255; Marmor (n 238) 55.

²⁴³ Dworkin (n 226) 255-256.

²⁴⁴ *ibid* 225.

²⁴⁵ *ibid* 66; Marmor (n 238) 55.

(i) Attacking Hercules' Task: The Problem of Individuation of Laws

In the above example, Hercules, when constructing a scheme of principles, justifies the fair comment defence by stating that people are entitled to exercise their freedom of speech to make comments on matters of public interest. Is this statement of principle, however, really a “principle” in the sense Dworkin has in mind? Dworkin’s discussion of principles fails to take into account what Joseph Raz calls, the problem of individuation of laws.²⁴⁶ The doctrine of individuation tells us that there is a fundamental distinction between a statement of principle that is a statement of the content of one complete law and another that is an abbreviated reference to a number of laws.²⁴⁷ Let us consider its application to Dworkin’s theory by way of two scenarios:

Scenario 1: Hercules’ jurisdiction recognises the principle of freedom of speech because of the enactment of a law expressly enshrining and preserving freedom of speech without any limitation.

Scenario 2: Hercules’ jurisdiction allows a high degree of speech apart from limitations in laws in relation to, for example, national security.

Only in scenario 1 may Hercules justify the fair comment defence by stating that people are entitled to exercise their freedom of speech to make comments on matters of public interest. Here, Hercules’ statement is a statement of the content of one particular law and a principle in the Dworkinian sense, i.e. one that “point[s] [the court] to particular decisions about legal obligation in particular circumstances”.²⁴⁸ In scenario 2, Hercules’ statement, despite taking the form of a principle, is only an abridged reference to a number of laws. It is not a statement of the content of a single law and thus not a Dworkinian principle.²⁴⁹ Put simply, therefore, not everything that

²⁴⁶ Joseph Raz, ‘Legal Principles and the Limits of Law’ (1972) 81 YLJ 823, 828.

²⁴⁷ *ibid* 829.

²⁴⁸ Dworkin (n 230) 24.

²⁴⁹ Raz (n 246) 828.

looks like a principle is a principle in the Dworkinian sense. What does this mean for Hercules' interpretive task? Dworkin's claim is *empirical*. He thinks that real judges decide hard cases by reference to principles in his sense of the term. Hercules, in constructing a scheme of principles, is meant to illustrate this.²⁵⁰ Two lessons may be drawn from applying the doctrine of individuation to Dworkin's discussion of principles.

First, judges do not, in fact, decide hard cases by reference to principles in the Dworkinian sense. Say, for example, we were to now put before Hercules a fictitious common law case involving a defendant who made defamatory comments about the private life of a public official. Hercules sets out the following three candidates for the best interpretation of the precedent cases:

- (1) people are entitled to exercise their freedom of speech to make comments on matters of public interest;
- (2) people have a moral right to compensation for defamatory comments made against them; and
- (3) people have a moral right to compensation when their right to private life is infringed.²⁵¹

If Hercules settles on (1), he must apply the fair comment defence and decide for the defendant; if on (2) or (3), for the claimant. But regardless of which Hercules chooses, we now know that (1) and/or (2) and/or (3) may not actually be principles in the Dworkinian sense, depending on how they are incorporated by Hercules' jurisdiction. It follows that, since Dworkin's discussion of principles does not account for the problem of individuation, Hercules (or any judge, for that matter) may well decide the case by reference to a principle that is not, in fact, a principle in the Dworkinian sense. If this is correct, Hercules cannot claim that, in constructing a body of principles, he is doing explicitly what real judges do in a less

²⁵⁰ For this reason, JW Harris describes Hercules as "an embodiment of assumptions (inexpertly and often inarticulately) contained within the reasoning-processes of real live judges": Harris (n 241) 49.

²⁵¹ These interpretations are set out by way of example. In a real case, there is likely to be more than three interpretations of the precedent cases.

conscious manner. Hard cases, like the one we just considered, are not *always* decided by reference to Dworkinian principles.

Second, it follows that the apparent abundance of principles is deceptive.²⁵² The body of principles Hercules constructs will almost certainly contain statements of principles that are not statements of the content of one particular law, but brief allusions to a number of laws.

(ii) *Attacking Hercules' Method: The Problems of Indistinguishability and Incommensurability of Criteria*

As explained above, Hercules determines the best interpretation of the precedent cases on two criteria: fit and appeal. The former plays a dual role. At the interpretive stage, it is a threshold requirement.²⁵³ At the post-interpretive stage, it plays a more evaluative role.²⁵⁴ Hercules' convictions of fit, when operating as a threshold requirement, must be independent of his convictions about appeal, failing which the former will not constrain the latter.²⁵⁵ There however lies a problem here. The exercise of fit and the exercise of appeal both rest on Dworkin's concept of *community*. The former instructs Hercules to have regard to the legal history of his community, whereas the latter requires Hercules to seek an interpretation which "shows the community in a better light...from the standpoint of political morality".²⁵⁶ Because this common strand of community runs through both exercises, it is difficult to wholly isolate one from the other and, accordingly, for the criterion of fit to act as a threshold requirement.²⁵⁷ We refer to this as the problem of indistinguishability of criteria.

²⁵² Raz (n 246) 828.

²⁵³ Dworkin (n 226) 255.

²⁵⁴ *ibid* 66; Marmor (n 238) 55.

²⁵⁵ Dworkin (n 226) 67; Marmor (n 238) 62.

²⁵⁶ Dworkin (n 226) 249, 255.

²⁵⁷ Raymond Wacks, *Understanding Jurisprudence* (3rd edn, OUP 2012) 141; Nigel E Simmonds, 'Imperial Visions and Mundane Practices' (1987) 46 CLJ 465, 479.

More recently, Dworkin has stated that the two criteria are in fact “interdependent and *neither prior to the other*”.²⁵⁸ This statement makes an even bigger muddle of his theory of interpretation. If the criterion of fit is not prior to the criterion of appeal, it seems to follow that fit now does not operate as a threshold requirement and only operates in the post-interpretive stage. If correct, this is a substantial diminishment of its role. If incorrect, it is nevertheless clear that Dworkin has left unclear the purpose of fit. All in all, Hercules offers judges a criterion of fit which neither has a clear purpose nor is applicable without the engagement of their moral and political values.²⁵⁹

Putting aside the problem of indistinguishability of criteria, let us now ask: how does Hercules decide between competing interpretations of precedent cases? For Dworkin, the answer lies in “the general *balance* of political virtues” embodied in the competing interpretations.²⁶⁰ However, John Finnis aptly notes that in the absence of any metric to commensurate the criteria of fit and appeal, the instruction to “balance” means that Hercules does no more than bear in mind fit and appeal, and then *choose*.²⁶¹ Finnis calls this the problem of incommensurability of criteria. That is, the criteria of fit and appeal are incommensurable and therefore are impossible to quantify.²⁶²

How does the problem of incommensurability affect Hercules’ method of adjudication? Consider the fictitious case we put before Hercules earlier. Hercules will decide which one of the eligible interpretations – (1), (2) and (3) – is the best interpretation based the criteria of fit and appeal. It follows that depending on which criterion we consider, (1), (2) and (3) may be ranked in different orders.²⁶³ On the scale of fit, (1) may be better than (2); and (2) than (3). On the scale of appeal, (3) may be better than (2); and

²⁵⁸ Ronald Dworkin, ‘Ronald Dworkin Replies’ in Justine Burley (ed), *Dworkin and His Critics* (Blackwell 2004) 382 (emphasis added).

²⁵⁹ Wacks (n 257) 141.

²⁶⁰ Dworkin (n 226) 256 (emphasis added).

²⁶¹ John Finnis, ‘On Reason and Authority in Law’s Empire’ (1987) 6 *Law and Philosophy* 357, 374.

²⁶² Finnis (n 261) 373.

²⁶³ John Finnis, ‘Natural Law and Legal Reasoning’ (1990) 38 *Cleveland State Law Review* 1, 8.

(2) than (1). Dworkin has made abundantly clear that Hercules' decision will depend on how his convictions about fit and appeal contest with each other, and, in the latter dimension, how his convictions about fairness, justice and procedural due process contest with each other.²⁶⁴ But until Hercules is able to demonstrate how he quantifies these different criteria, these criteria are incommensurable, and there is, in reality, no sufficient reason to declare (1) a better interpretation than (3), or (3) than (1), or (2) than either.²⁶⁵ It appears that Hercules is merely making an evaluative *choice* between eligible interpretations, which is not guided by law as integrity but established by his sentiments and dispositions.²⁶⁶ If this is correct, Hercules cannot be a good model for judges to emulate in hard cases, wherein, by definition, the rank order of the eligible interpretations will always be different on the scales of fit and appeal.²⁶⁷ Methodologies and approaches can be emulated; sentiments and dispositions cannot.

Finnis goes further to imply that, after Hercules has made his choice, his answers will seemingly have a supremacy over the rejected alternative(s).²⁶⁸ The writer of this paper does not seek to push that far as Finnis is mistaken here. Dworkin has made clear that Hercules' answers do not define law as integrity as a conception of law.²⁶⁹ Law as integrity, instead, consists in his method. In recognition of this, this paper's critique focuses on Hercules' method, rather than his answers.

²⁶⁴ Dworkin (n 226) 410-411.

²⁶⁵ Finnis (n 263) 8.

²⁶⁶ *ibid* 9; Finnis (n 261) 374.

²⁶⁷ Finnis (n 263) 8.

²⁶⁸ *ibid* 9; Finnis (n 261) 374.

²⁶⁹ Dworkin (n 226) 239.

IV. Defending Hercules

In *Law's Empire*, Dworkin responds to what he terms “familiar objections” – a list of charges critics have pressed against Hercules.²⁷⁰ Only one attacks Hercules’ method of adjudication head-on: the charge that Hercules is substituting his own political judgment for the politically neutral interpretation of precedent cases, rendering his decisions more tenable for a politician than a judge. Dworkin, in his response, justifies Hercules’ reliance on justice and fairness in deciding which interpretation of precedent appeals most.²⁷¹ It, therefore, represents his strongest argument as to why Hercules is a good model for judges.

For Dworkin, the notion that Hercules’ interpretations must match the interpretations of past judges is by no means politically neutral.²⁷² It is based on certain political values which compete with the demands of justice and fairness. A judge, who strictly abides by the interpretations of past judges, thinks that these political values eliminate any competing demands of justice and fairness, and is therefore no less of a politician than Hercules.

Dworkin is however mistaken in assuming that the demands of justice and fairness *always* compete with these political values. It may be that the rules of Hercules’ jurisdiction – most likely constitutional in nature – have settled the question of how much the interpretations of past judges are to constrain him from satisfying the demands of justice and fairness, in which case, Hercules is opening up and unsettling questions which have already been resolved.²⁷³ Put simply, Dworkin is instructing Hercules to resolve questions which the law is meant to resolve.²⁷⁴

²⁷⁰ *ibid* 258-266, 397-399.

²⁷¹ *ibid* 259.

²⁷² *ibid* 260.

²⁷³ Scott J Shapiro, *Legality* (Harvard University Press 2011) 311; Wacks (n 257) 138.

²⁷⁴ Shapiro (n 273) 307.

V. Conclusion

A critic must not only raise Hercules' inadequacies but prove that they prevent him from providing a good model for judges. This writer has endeavoured to discharge this burden by arguing that two fundamental aspects of Hercules – his task of producing a theory of law and his method of adjudication – are plagued by the problem of individuation of laws, and the problems of indistinguishability and incommensurability of criteria respectively. It has been illustrated how Hercules' interpretive task does not reflect what real judges do and why real judges cannot and should not emulate his method of adjudication. Finally, it has been argued that Hercules opens up moral questions which his jurisdiction may well have resolved. For these reasons, it is concluded that Hercules does not provide a good model for real judges to emulate.

Pregnant women may have moral obligations to foetuses they have chosen to carry to term, but the law should never intervene in a woman's choices during pregnancy

*Elizabeth Chloe Romanis*²⁷⁵

For the duration of gestation, the welfare of a foetus is dependent on the behaviour of the woman carrying it. Pregnant women who make detrimental lifestyle choices risk severely harming the foetus and condemning it to substance addiction, long-term developmental impairments, severe birth deformities, or premature death. It will be assumed that once a woman decides to carry her pregnancy to term she is morally responsible for that choice, and consequently for foetal welfare. It does not follow, however, that the law should place obligations on pregnant women that mirror these responsibilities. This paper explores the extent to which civil and criminal law enforces these moral responsibilities and by actively intervening in choices during pregnancy and childbirth. It then explores the extent to which any imposition of liability or intervention can be justified, with a focus on foetal welfare and the privacy, autonomy, and liberty of women. It argues that no active or passive intervention with the decision-making of pregnant women is justifiable – even if their behaviour risks severely harming an unborn child – because of the constraints it would place on the freedom of all women and because intervention is unlikely to secure better foetal welfare.

²⁷⁵ LL.M. Candidate, The University of Manchester, School of Law.

I. Introduction

Pregnant women who drink alcohol or take illegal drugs can give birth prematurely²⁷⁶ to children with severe deformities²⁷⁷ and substance addictions, causing long-term problems with their development.²⁷⁸ Complications during childbirth can become a serious threat to the foetus if the mother does not undertake medical procedures to minimise such risks. However, foetal welfare can directly conflict with the interests or preferences of the mother: pregnant women may have to make drastic lifestyle changes or be subjected to bodily interference to protect the foetus. Such conflicts are becoming more frequent as advances in medical science^{279,280} highlight more potential causes of harm in utero,²⁸¹ and as political pressure mounts to solve the problems states face in caring for children who suffered avoidable injury in utero.

A woman can terminate a pregnancy until the 24th week of gestation if continuation of pregnancy would involve greater risk than termination.²⁸² When considering risk, regard must be given to the pregnant woman's actual or reasonably foreseeable environment.²⁸³ Statistics suggest that the risk of abortion is less than the risk of childbirth before the 12th week of gestation.²⁸⁴ After the 24th week, by contrast, termination is legal only if there is a risk of grave injury to,²⁸⁵ or death²⁸⁶ of, the woman, or if the foetus is disabled.²⁸⁷

²⁷⁶ Carol Simon, 'Prenatal Liability for Prenatal Injury' (1978) 14 Columbia Journal of Law and Social Problems 47, 74.

²⁷⁷ *ibid.*, 73-74.

²⁷⁸ Margery W Shaw, 'Conditional Prospective Rights of the Foetus' (1984) 5 Journal of Legal Medicine 63, 73.

²⁷⁹ E.g. James Gallagher, 'Mother's Diet During Pregnancy Alters Baby's DNA' *BBC News* (London 18 April 2011) <<http://www.bbc.co.uk/news/health-13119545>> accessed 04 January 2016.

²⁸⁰ Bartha M Knoppers, 'Picard Lecture in Health Law – 1992: Human Genetics: Parental, Professional and Political Responsibility,' (1993) 1 Health Law Journal 13, 15.

²⁸¹ Laura M Purdy, 'Are Pregnant Women Fetal Containers?' (1990) 4 Bioethics 273, 275.

²⁸² Abortion Act 1967, s.1 (1) (a).

²⁸³ *ibid.*, s.1 (2).

²⁸⁴ Margaret Brazier and Emma Cave, *Medicine Patients and the Law*, (5th edn, Penguin Books, London 2011), 390.

²⁸⁵ Abortion Act 1967, s.1 (1) (b).

²⁸⁶ *ibid.*, s.1 (1) (c).

²⁸⁷ *ibid.*, s.1 (1) (d).

This paper will assume that once a woman has decided to carry her pregnancy to term she is morally responsible for that choice. A responsible pregnant woman always “consider[s] the best interests of the unborn [child]...[she] has decided to mother”²⁸⁸ because it is dependent on her.²⁸⁹ She “must avoid harmful substances, act to meet [foetal] needs... at times submit to bodily invasions”,²⁹⁰ avoid unhealthy environments, engage in healthy activities, and make sensible medical decisions.²⁹¹ Moral and legal responsibilities can overlap, yet not every moral responsibility is translated into law.²⁹² Law ought only to enforce moral responsibilities to meet a legitimate aim or when it effectively deters significant harm.

This essay will examine whether the law recognises pregnant women’s moral responsibilities and intervenes in choices during pregnancy and childbirth. It will demonstrate that there are inconsistencies in the law’s approach to harmful choices during pregnancy, and that intervention is unjustified because it infringes upon women’s basic freedoms and does not better secure foetal welfare.

²⁸⁸ Heather Draper, ‘Women, Forced Caesareans and Antenatal Responsibilities’ (1996) 22 *Journal of Medical Ethics* 327, 329.

²⁸⁹ Margaret Brazier, ‘Liberty, Responsibility, Maternity’ (1999) 52 *Current Legal Problems* 359, 367.

²⁹⁰ Laura M Purdy (n 281), 275.

²⁹¹ *ibid.*, 275.

²⁹² Peter Cane, *Responsibility in Law and Morality*, (1st edn, Hart Publishing, Oxford 2002), 1.

II. Legal status of the foetus

A foetus, with no existence independent of its mother, has no legal personality²⁹³ in England and Wales. It has no independent rights that can be enforced against another²⁹⁴ until it has been born, at which point it “attains the status of a legal persona”.²⁹⁵ The conception of birth as the point at which rights are ascertained is so pervasive in English law it has been described as ‘unassailable’.²⁹⁶ This is consistent with the position of the European Court of Human Rights, which has held that a foetus does not have an unqualified right to life²⁹⁷ because it cannot be separated from its mother. Her rights must be prioritised over those of an inchoate person.²⁹⁸

Although the law is clear that a foetus is not equal to a person, it still affords it some protection. The Infant Life (Preservation) Act 1929 criminalises child destruction – when a person, with intent and by wilful act, destroys the life of a child capable of being born alive.²⁹⁹ After the 24th week of gestation, there is prima facie proof that the foetus is viable.³⁰⁰ The Abortion Act 1967 provides that medical termination of pregnancy is a defence to child destruction.³⁰¹ By limiting the grounds on which a woman can receive a medical termination to more extreme circumstances, after the 24th week of gestation,³⁰² the Act appears to protect a ‘right to be gestated’, though subordinate to maternal interests in health and life.³⁰³ These provisions do not award the foetus legal personality, nor do they protect the foetus from harm caused by the mother’s choices during pregnancy and childbirth³⁰⁴ not explicitly intended to harm the foetus.

²⁹³ *Paton v British Pregnancy Advisory Service Trustees* [1979] QB 276.

²⁹⁴ *C v S* [1987] 1 All ER 1230.

²⁹⁵ *Re MB (An adult: medical treatment)* [1997] 8 Med LR 217; *Paton* (n 293).

²⁹⁶ Ian Kennedy and Andrew Grubb, *Medical Law: Text with Materials*, (3rd edn, Butterworths, London 2000), 1487.

²⁹⁷ European Convention of Human Rights, Article 2.

²⁹⁸ *X v UK* 7215/75 (1981) ECHR 6.

²⁹⁹ Infant Life (Preservation) Act 1929, s.1 (1).

³⁰⁰ *ibid*, s.1 (2).

³⁰¹ Abortion Act 1967, s.1.

³⁰² *ibid*, s.1 (1) (b) – (d).

³⁰³ Gerard Casey, ‘Pregnant Women and Unborn Child: Legal Adversaries?’ (2002) 8 *Medico-Legal Journal of Ireland* 75, 83.

³⁰⁴ Margaret Brazier (n 289), 364.

III. Liability

Liability might prevent foetal harm by retrospectively sanctioning women who make harmful choices during pregnancy.³⁰⁵ This section will examine the extent to which liability exists in law and whether, on the one hand, it is a legitimate infringement on women's rights, and on the other, effectively secures foetal welfare.

1) *Civil liability*

A child, born alive,³⁰⁶ can pursue an action in negligence for damage in utero because at birth they attain legal personality and inherit "the damaged body for which the [defendant] (on the assumed facts) [is] responsible".³⁰⁷ The Congenital Disabilities (Civil Liability) Act 1976 provides that 'the child's own mother' cannot be held liable for antenatal conduct that results in disability when born alive.³⁰⁸ The same exception is not made for fathers. No matter how negligent a mother is during pregnancy,³⁰⁹ and however serious the damage accrued, the child has no recourse, with one exception: a woman may be held liable for negligent driving whilst pregnant.³¹⁰ This is justifiable because mandatory³¹¹ insurance policies³¹² spread the cost of compensation and drivers are already under a duty to take reasonable care on the road.³¹³

It is easy to establish a duty of a mother towards her foetus under the 'neighbour principle'.³¹⁴ However, establishing breach, another element of the tort of negligence,³¹⁵ would result in a disproportionate restriction of liberty. That is, the reasonable

³⁰⁵ Andrea Evenski, 'The Use of Incarceration and Forced Detention as a Method of Fetal Protection for Substance-abusing Pregnant Women,' (2001) 58 *Journal of Current Surgery* 402, 402.

³⁰⁶ Congenital Disabilities (Civil Liability) Act 1976, s.4 (2).

³⁰⁷ *Burton v Islington Health Authority* [1993] QB 204, 219 (Dillon LJ).

³⁰⁸ Congenital Disabilities (Civil Liability) Act 1976, s.1 (1).

³⁰⁹ Margaret Brazier and Emma Cave (n 284), 316.

³¹⁰ Congenital Disabilities (Civil Liability) Act 1976, s.2.

³¹¹ Road Traffic Act 1988, s.143 (1) (a).

³¹² J Mason and G Laurie, *Mason and McCall Smith's: Law and Medical Ethics*, (8th edn, Oxford University Press, Oxford 2011), 394.

³¹³ Congenital Disabilities (Civil Liability) Act 1976, s.2.

³¹⁴ *Donoghue v Stevenson* [1932] UKHL 100.

³¹⁵ J Mason and G Laurie (n 312), 142.

standard of care in pregnancy, which the negligent pregnant woman falls below, would be hard to establish³¹⁶ and make equitable.³¹⁷ Some lifestyle habits are known to harm the foetus, yet the level at which harm accrues is medically disputed³¹⁸ and contradictory advice is given to women.³¹⁹ Following authoritative yet contradictory advice, pregnant women might make choices that the law may regard as insufficient. This would prejudice the most vulnerable, who may be less able to understand and make sense of such advice; it may also prejudice those who struggle with poverty³²⁰ or addiction,³²¹ factors which might prevent them from meeting the standard of the ‘ordinary pregnant woman’. Setting a reasonable standard of behaviour during pregnancy places such women in difficult situations when, for example, they are aware that sudden substance withdrawal might harm themselves and their baby,³²² but continued drug use will harm their baby. Notwithstanding this problem, having different standards of reasonable behaviour would result in inconsistencies in law that might lead to further confusion and complexity.

The liberty of women, in general, is restricted earlier in pregnancy, potentially before a woman knows she is pregnant;³²³ and even before conception,³²⁴ the foetus can be harmed. Liability during pregnancy demands that “fertile, sexually active women of childbearing age should act at all times as if they were pregnant”³²⁵ and live restricted by the standards of the reasonable pregnant woman. Furthermore, women found in breach of this standard are

³¹⁶ Margaret Brazier, ‘Parental Responsibilities, Foetal Welfare and Children’s Health’ in Caroline Bridge (ed.), *Family Law: Towards the Millennium: Essays for PM Bromley*, (1st edn, Butterworths London 1997), 272.

³¹⁷ Margaret Brazier and Emma Cave (n 284), 334.

³¹⁸ Margaret Brazier (n 316), 272.

³¹⁹ Margaret Brazier and Emma Cave (n 284), 334.

³²⁰ *ibid.*, 334.

³²¹ Margaret Brazier (n 289), 379.

³²² Margaret Brazier (n 316), 272.

³²³ Margaret Brazier and Emma Cave (n 284), 334.

³²⁴ H Briggs, ‘Pre-Pregnancy Diet Permanently Influences Baby’s DNA’ *BBC News* (London 30 April 2014) <<http://www.bbc.co.uk/news/health-27211153>> accessed 08 January 2016.

³²⁵ Sheila Mclean, *Old Law, New Medicine: Medical Ethics and Human Rights*, (1st edn, Pandora, London 1999), 66.

likely to be those that have not complied with medical advice, regardless of intent or reason.³²⁶

On the other hand, there may be some cases, such as those involving heavy drug abuse, where establishing negligence would be at little cost to the liberty of women in general. In these situations, negligence is apparent but responsibility is precluded, as per the 1976 Act. Thus, the important question is: ought we to allow damage in these circumstances to go without remedy?

The Law Commission has posited that pregnant women ought not to be liable because they are less likely to be able to afford the compensation awarded, it would disturb the parental bond and may be used by a father as a weapon in a familial dispute.³²⁷ Brazier and Cave, however, believe that reasons related to the financial position of women are out-dated.³²⁸ Women are now, to some degree, more financially empowered, yet female-headed single-parent families remain the poorest in the UK and many live below the poverty line.³²⁹ If a child is removed from its mother and compensation is awarded, there will be little chance of recovery and the burden might place vulnerable mothers in more dire circumstances. Nevertheless, this does not prove that such behaviour is less culpable; it merely explains limitations that must be placed on compensation awarded. Kennedy and Edwards note that arguments for the maternal exception related to familial bonds and disputes apply equally to fathers.³³⁰ This reasoning is not the most important to the exception because in these situations most children born alive will be removed from their mothers.³³¹

A child can claim against their mother for harm accrued just after birth but cannot claim for harm in utero. This is concerning because the mother has the potential to cause more harm during

³²⁶ Margaret Brazier (n 316), 274.

³²⁷ Law Com Report No 60, *Injuries to Unborn Children* (1974, HMSO).

³²⁸ Margaret Brazier and Emma Cave (n 284), 334.

³²⁹ Sally Sheldon, 'Unwilling Fathers and Abortion: Terminating Men's Child Support Obligations?' (2003) 66 *Modern Law Review* 175, 188.

³³⁰ Ian Kennedy and R Edwards, 'A Critique of the Law Commission Report on Injuries to Unborn Children and the Proposed Congenital Disabilities (Civil Liability) Bill' (1975) 1 *Journal of Medical Ethics* 116, 119.

³³¹ E.g. *D v Berkshire County Council* [1987] 1 All ER 20.

gestation than after birth.³³² The conduct of the mother poses a greater risk to the foetus³³³ than any other person who might be held liable because it is affected by everything she does. This results in extreme inequity when children are not compensated.³³⁴ While it is regretful that some children receive no compensation for injury, civil liability causes much more significant harm to the mother's liberty and stigmatises women limited by their circumstances. Moreover, it is unlikely that foetuses would be better protected from avoidable injury if women are incentivised when they cannot control their addiction or circumstances. Even if the risk of mothers pursuing illegal terminations, causing serious personal harm and harm to surviving children is small, the damage may be so significant that the law should not incentivise its occurrence.

2) Criminal liability

In *Attorney-General's Reference (No 3 of 1994)*, it was held that a person is guilty of manslaughter if they undertake dangerous or unlawful action causing injury to a foetus that is subsequently born alive before dying. A deliberate attempt to harm a foetus that is born alive before dying could amount to murder.³³⁵ Lord Mustill, *obiter*, posited that harm short of death might give rise to criminal liability.³³⁶ The child must be born alive before dying for a homicide offence to be committed because homicide can only be committed against persons.³³⁷ The reasoning did not mention but does not exclude, pregnant women from criminal liability. Women may be guilty, when their child is born alive and subsequently dies, of unlawful act manslaughter because of drug abuse, or of gross negligence manslaughter because of alcohol consumption.³³⁸ These prosecutions have not yet been tested.

³³² P Pace, 'Civil Liability for Pre-Natal Injuries' (1977) 40 Modern Law Review 141, 157.

³³³ Margaret Brazier (n 316), 266.

³³⁴ P J Pace, 'Civil Liability for Pre-Natal Injuries' (1977) 40 Modern Law Review 141, 157.

³³⁵ *Attorney-General's Reference (No 3 of 1994)* [1997] 3 All ER 936.

³³⁶ *ibid*, 949 (Lord Mustill).

³³⁷ Edward Coke, *The Third Part of the Institutes of the Laws of England*, (M Flesher for W Lee and D Pakeman, London 1644), 47.

³³⁸ Margaret Brazier and Emma Cave (n 284), 336.

Pickworth argues that this “raises the possibility that English law will follow that of the United States in extending criminalization” to those who cause harm short of death to a foetus.³³⁹ This logical development was not precluded³⁴⁰ and forced caesarean cases to allude to a pattern of willingness to condemn poor antenatal conduct.³⁴¹ Limited protection for the foetus in civil law³⁴² might encourage the criminal law to intervene³⁴³ to prioritise foetal welfare over maternal autonomy.³⁴⁴ Lord Mustill’s *obiter* comments and Pickworth’s argumentation regarding the potential of women to be liable for harm less than death has, at least for now, been put to bed.

The Court of Appeal in *CP v CICA*,³⁴⁵ held that a child born alive with foetal alcohol syndrome due to alcohol consumption in pregnancy was not a victim of a crime. CP alleged that excessive alcohol consumption amounted to the malicious administration of poisoning to inflict grievous bodily harm (GBH). The offence is committed when malicious administration of any noxious substance causes harm to “any other person”.³⁴⁶ The offence was not committed because the foetus was not a person at the time of the relevant act.³⁴⁷ The Court maintained that a mother abusing alcohol could be guilty of manslaughter if the foetus dies after being alive, but not GBH because a time lag between the infliction of fatal injury and death does not preclude homicide, whereas GBH is committed at the point of violence.³⁴⁸ When a poisonous substance is administered in utero, GBH is inflicted on the foetus, not a person. The outcome in *CP* was determined by a literal interpretation of

³³⁹ Emma Pickworth, ‘Substance Abuse in Pregnancy and the Child Born Alive’ (1998) 27 *Anglo-American Law Review* 472, 472.

³⁴⁰ *ibid.*, 475.

³⁴¹ *ibid.*, 476.

³⁴² *Re F (In Utero)* [1988] Fam. 122.

³⁴³ Emma Cave, *The Mother of All Crimes: Human Rights, Criminalization and the Child Born Alive*, (1st edn, Ashgate Publishing, London 2004), 29.

³⁴⁴ Sara Fovargue and José Miola, ‘Policing Pregnancy: Implications of the Attorney-General’s Reference (No 3 of 1994)’ (1998) 6 *Medical Law Review* 265, 266.

³⁴⁵ *CP (A Child) v Criminal Injuries Compensation Authority* [2015] QB 459.

³⁴⁶ Offences Against the Person Act 1861, s.23.

³⁴⁷ *CP v CICA* (n 345), 471 (Treacy LJ).

³⁴⁸ *ibid.*, 472 (Treacy LJ).

statute.³⁴⁹ Thus, the Court was constricted, to avoid overriding Parliament, by the formulation of the statutory offence.

Lord Dyson MR cautioned that “the court should be slow to interpret general criminal legislation as applying” in the context of harm to fetuses.³⁵⁰ Parliament may be willing to entertain a change in the law because of the inconsistencies that have arisen in this area. For example, women may be criminally liable when their child, born alive, dies because of harm they caused, but not when it survives with serious injury despite the consequences arguably being much graver. Should women be criminally liable for damage to fetuses?

Holding pregnant women criminally liable is a constraint on their autonomy, liberty, and privacy. Imposing a criminal standard dependent on the condition of pregnancy, by outlawing behaviour that is legal but for pregnancy, results in discrimination,³⁵¹ inconsistency, and vagueness in the law.³⁵² It disproportionately punishes those already ostracised from society by factors often beyond their control, such as addiction and poverty. Limiting a woman’s rights to autonomy, liberty, and privacy cannot be tolerated without a reasonable belief that it would immeasurably benefit others.³⁵³ A foetus is not a legal person, so to curtail pregnant women’s liberty for their benefit is a significant change in law that would undoubtedly affect abortion law and restrict women’s ability to make choices during pregnancy. It is unlikely that retrospective criminal sanctions would improve foetal welfare. It would not deter addictive behaviours but instead, introduces an incentive to abort any child³⁵⁴ to avoid punishment. Indeed, women are likely to evade medical treatment³⁵⁵ and social care, for fear of reprise from professionals, with wide ramifications for the welfare of any children born.

³⁴⁹ Caroline Bridge, ‘Crime’ (2015) 45 Family Law 123, 124.

³⁵⁰ *CP v CICA* (n 345), 475 (Lord Dyson MR).

³⁵¹ Emma Cave (n 343), 94.

³⁵² *ibid*, 93.

³⁵³ Margaret Brazier (n 316), 293.

³⁵⁴ *ibid*, 298.

³⁵⁵ *ibid*, 290.

IV. Direct intervention

This section addresses the questions: a) when the judiciary consider active intervention in the choices of pregnant women, are they willing to supersede autonomy? and b) does this better secure foetal welfare?

1) *Protective custody*

In the United States, pregnant women have been convicted of prenatal child abuse and incarcerated to protect the foetus.³⁵⁶ One woman, for example, was subjected to a protective custody order to protect her foetus, although this was reversed on appeal.³⁵⁷ In England and Wales, courts can exercise their jurisdiction to make a vulnerable child a ward of court and then must ensure, in judgments, that the “paramount consideration” is the ward’s welfare.³⁵⁸ It is not possible, following *Re F*,³⁵⁹ to make a foetus a ward of court despite acknowledgement that a need for this power exists.³⁶⁰ The Canadian Supreme Court persuasively affirmed this judgment in 1997.³⁶¹ In *Re F*, the local authority sought to ensure that a mother remained in hospital until she had given birth to protect the foetus. The Court of Appeal refused to extend its wardship jurisdiction for several reasons.

Firstly, because the foetus has no legal personality³⁶² and thus is not a minor,³⁶³ the court could not exercise jurisdiction without misconstruing statutory provisions.³⁶⁴ A change in law with regard to foetal wardship was stipulated to be a matter for Parliament.³⁶⁵ However, in *D v Berkshire*,³⁶⁶ the House of Lords acknowledged that a woman’s conduct during pregnancy was

³⁵⁶ *Reinesto v Arizona* 894 P.2d 733 (Wise Ct App 1995).

³⁵⁷ *Wisconsin v Angela M.W. v Kruzicki* 541 N.W. 2d 482 (Wise Ct App 1995).

³⁵⁸ *Re E (A Minor) (Wardship: Court’s Duty)* [1984] 1 WLR 156, 159 (Lord Scarman).

³⁵⁹ *Re F* (n 342).

³⁶⁰ *ibid*, 145 (Staughton LJ).

³⁶¹ *Winnipeg Child and Family Services (Northwest Area) v G 3* BHRC 611.

³⁶² *Paton* (n 293); *C v S* (n 294).

³⁶³ Family Law Reform Act 1969, s.1.

³⁶⁴ Senior Courts Act 1981, s.41 (1).

³⁶⁵ *Re F* (n 342) (May LJ), 144 (Balcombe LJ).

³⁶⁶ *D v Berkshire* (n 331).

admissible evidence in care proceedings and that drug abuse during pregnancy demonstrated that the child's development was prevented or neglected. This was an acknowledgement that antenatal behaviour ought not to be ignored and that the law sometimes does concern itself with the treatment of the foetus.³⁶⁷ The case resulted in fear that the judiciary might yet intervene by extending its wardship jurisdiction, although no such intervention has materialised.³⁶⁸

Secondly, because there is an inevitable conflict between the foetus's interests and the mother's liberties and, if deemed a ward, the foetus's interests would be paramount,³⁶⁹ an intolerable³⁷⁰ sacrifice of the mother's rights^{371,372} would occur. Balcombe LJ has commented that "since an unborn child has...no existence independent of its mother, the only purpose of extending the jurisdiction...is to enable the mother's actions to be controlled".³⁷³ Moreover, "the court is in no position...to decide in what circumstances and with what safeguards there should be power to restrict the liberty of the mother" to prevent foetal harm.³⁷⁴ Thirdly, because orders made following wardship would be difficult to enforce if the mother was non-compliant, it was considered inappropriate for compliance to be compelled by force, and impossible "to consider with any equanimity that the Court should seek to enforce an order by committal".³⁷⁵

The judgment is thorough in explaining that the rights of women cannot be subordinated to protect a foetus. It negates the question of foetal welfare because the foetus is not a legal person, consistent with the wider civil law. If Parliament, however, were to consider a change in the law, then the question of how the courts' wardship jurisdiction would affect foetal welfare becomes relevant.

³⁶⁷ Jane Fontin, 'Legal Protection for the Unborn Child' (1988) 51 *Modern Law Review* 54, 54.

³⁶⁸ Margaret Brazier (n 316), 281.

³⁶⁹ Children Act 1989, s.1 (1).

³⁷⁰ *Re F* (n 342), 138 (May LJ).

³⁷¹ *Re F* (n 342), 135 (May LJ).

³⁷² N Lowe, 'Wardship and Abortion Prevention – Further Observations' (1980) 96 *Law Quarterly Review* 29, 30.

³⁷³ *ibid*, 143 (Balcombe LJ).

³⁷⁴ *ibid*, 145 (Staughton LJ).

³⁷⁵ *ibid*, 138 (May LJ).

To enforce wardship, a pregnant woman is likely to be detained. Detention to protect the foetus from harmful conduct was the motivation behind several decisions in the US.³⁷⁶ When a pregnant woman is detained, it allows the state to guarantee the safety of her foetus because access to harmful substances can be monitored and prevented.³⁷⁷ Glaze contends that detention is no different from quarantine, in which a person is detained for the benefit of third parties;³⁷⁸ here, this includes parents and society.³⁷⁹ However, this cannot be endorsed if benefits do not materialise.

Evenski argues that detention is likely to cause more harm to the foetus. Women with addictions become less likely to trust their physician³⁸⁰ and social services when they fear a breach of privacy.³⁸¹ This would deter the majority of affected women from seeking prenatal care “during which the physician would be able to educate [and] treat and counsel effectively on the effects of her [behaviour]”.³⁸² Even if a mother does not cease using harmful substances during pregnancy, access to prenatal care still increases the likelihood of better perinatal outcomes.³⁸³ Parliament should be wary of extending the wardship jurisdiction to avoid subjecting pregnant women to an instrument of absolute control that does not, *per se*, better secure foetal welfare.

³⁷⁶ E.g. *Wisconsin v Angela* (n 357).

³⁷⁷ Cynthia Glaze, ‘Combating Prenatal Substance Abuse: The State’s Current Approach and the Novel Approach of the Court-Ordered Protective Custody of the Fetus’ (1997) 80 *Marquette Law Review* 793, 813.

³⁷⁸ *ibid*, 812.

³⁷⁹ Denise Cahalane, ‘Court-Ordered Confinement of Pregnant Women’ (1989) 15 *New England Journal on Criminal and Civil Confinement* 203, 216.

³⁸⁰ Andrea Evenski (n 305).

³⁸¹ *ibid*, 403.

³⁸² *ibid*, 402.

³⁸³ N Milliken, ‘Maternal-Fetal Relationships: Medical Aspects’ in W Reich, *Encyclopedia of Bioethics Vol.3*, (1st edn, Macmillan, New York 1995), 1403-1407.

2) *Compelling caesareans*

Caesareans are sometimes necessary to mitigate a serious threat to the foetus. When these situations materialise, but the caesarean is refused, the most common legal challenge is that the woman lacks capacity under the Mental Capacity Act 2005,³⁸⁴ thus refusal is not autonomous³⁸⁵ and the procedure can be compelled in her best interests.³⁸⁶ In *Re MB*³⁸⁷ and *St George's v S*,³⁸⁸ the Court of Appeal was equivocal that caesareans are not legal when a patient with capacity refuses because “while pregnancy increases the personal responsibilities of the woman it does not diminish her entitlement” to refuse treatment.³⁸⁹ This is consistent with *Re F*.³⁹⁰ Despite this, the judiciary can interfere with a woman’s choices in childbirth because “the right to refuse...is frequently undermined by the courts’ use of available common law resources in determining capacity”.³⁹¹

In *Rochdale Healthcare Trust v C*, C was found to lack capacity because of the pains and stresses of labour.³⁹² This challenges the capacity of all pregnant women as it implies that, by virtue of their condition,³⁹³ they are incapable of decision-making.³⁹⁴ *Re MB*³⁹⁵ determined that phobia removed the ability to make decisions, but there was little explanation as to how.³⁹⁶ This precedent allows judges to “err on the side of finding incompetence” and determine that common and temporary fears render a pregnant

³⁸⁴ Mental Capacity Act 2005, s.3 (1).

³⁸⁵ *Re T (Adult: Refusal of Medical Treatment)* [1993] Fam 95, 117 (Butler Sloss LJ); *Re MB* (n 295)

³⁸⁶ Mental Capacity Act 2005, s.1 (5).

³⁸⁷ *Re MB* (n 295).

³⁸⁸ *St George's Healthcare NHS Trust v S* [1998] 3 WLR 936.

³⁸⁹ *ibid*, 957 (Judge LJ).

³⁹⁰ *Re F* (n 342).

³⁹¹ John Harrington, ‘Privileging the Medical Norm: Liberalism, Self-Determination and Refusal of Treatment’ [1996] 16 Legal Studies 248, 358.

³⁹² *Rochdale Healthcare (NHS Trust) v C* [1997] 1 FCR 274, 275 (Johnson J).

³⁹³ Sara Fovargue and José Miola (n 344), 287.

³⁹⁴ Jonathan Herring, *Medical Law and Ethics*, (4th edn, Oxford University Press, Oxford 2012), 345.

³⁹⁵ *Re MB* (n 295).

³⁹⁶ Robert Francis, ‘Compulsory Caesarean Sections: An English Perspective’ [1997-1998] 14 Journal of Contemporary Health Law and Policy 365, 388.

woman incapacitated.³⁹⁷ In *Re AA*,³⁹⁸ the judgment spent little time explaining its application of the capacity test to a schizophrenic woman.³⁹⁹ In *Royal Free NHS Trust v AB*,⁴⁰⁰ Hayden J explained his conclusion of incapacity, but appeared to suggest that the relevant standard of decision-making was rationality rather than capacity, when remarking that AB would have consented to the procedure were she not mentally ill because she was intelligent.⁴⁰¹ In *NHS Trust v FG*,⁴⁰² guidelines were suggested explaining how caesarean cases ought to be managed, including thorough capacity tests and insisting earlier applications should be made if problems later in pregnancy are anticipated. Keehan J, however, still spent little time explaining his finding of incapacity.⁴⁰³ There is inconsistency in the approach to, and potentially misapplication of, the capacity test that presently allows the judiciary to intervene with women's choices during pregnancy.

Forced caesareans necessitate the use of force that could not be contemplated in *Re F*.⁴⁰⁴ Why is it that forcing a woman to subvert her autonomy and liberty for a better foetal outcome in this context is less repulsive to the judiciary? Perhaps this is because compulsory caesareans are shorter interventions than those possible if a foetus was a ward of court. Yet, this is not a satisfactory explanation, due to the incredibly harmful nature of a forced surgery and its lasting effects.

Thorpe J, writing extra-judicially, speculated that "it is, perhaps, easier for an appellate court to discern principle than it is for a trial court to apply it in the face of judicial instinct...applications...are confined to judges of the Family Division...dedicated to upholding child welfare. It is simply unrealistic to suppose that the preservation of each life..." will not

³⁹⁷ *ibid*, 388.

³⁹⁸ *Re AA (Compulsorily Detained Patient: Elective Caesarean)* [2012] EWHC 4378.

³⁹⁹ Sara Fovargue and José Miola, 'Are we Still Policing Pregnancy?' in C Stanton, S Devaney, A Farrell and A Mullock (eds.), *Pioneering Healthcare Law: Essays in Honour of Margret Brazier*, (1st edn, Routledge, New York 2016), 249.

⁴⁰⁰ *Royal Free NHS Foundation Trust v AB* [2014] EWCOP 50.

⁴⁰¹ *ibid* [15] (Hayden J).

⁴⁰² *NHS Trust 1, NHS Trust 2 v FG* [2014] EWCOP 30.

⁴⁰³ Sara Fovargue and José Miola (n 399), 253.

⁴⁰⁴ *Re F* (n 342), 138 (May LJ).

concern the judge.⁴⁰⁵ In the Court of Appeal, it was also evident that the judiciary regretted that the legal status of the foetus meant there was no protection in this situation. In *St George's v S*, it was noted that “whatever else it may be, a 36-week foetus is not nothing; if viable, it is not lifeless and it is certainly human”.⁴⁰⁶ Whilst the foetus’s interests are of no concern in determining the outcome of a caesarean case,⁴⁰⁷ the judiciary find it difficult to ignore. In *Re AA*,⁴⁰⁸ Mostyn J noted “...harsh though it is, the interests of this unborn child are not the concern of this court...”⁴⁰⁹ and in *Norfolk and Norwich v W*,⁴¹⁰ Johnson J was clear that the woman’s interests were paramount but “the reality was that the fetus was a fully formed child, capable of a normal life if only it could be delivered from the mother”.⁴¹¹

Forced caesareans can only be performed when the patient lacks capacity. Judges appear to be succumbing to the temptation to rule ‘on the safe side’ with capacity, protecting the foetus by authorising caesareans especially in time-pressured situations with less thorough deliberation.⁴¹² These cases demonstrate a trend of interference “at odds with both the criminal and civil law”,⁴¹³ with drastic implications for maternal liberty and foetal welfare. The capacity test should be properly applied to ensure female autonomy is not being subverted to allow judicial intervention. We can see in recent cases⁴¹⁴ that some pregnancies are continuing to be ‘policed’.⁴¹⁵

The ability to court-order compliance with procedures in childbirth terrifies vulnerable women,⁴¹⁶ and can affect the doctor-patient relationship because doctors can now “abandon the language

⁴⁰⁵ Thorpe LJ, ‘The Caesarean Section Debate’ [1997] 27 Family Law 663, 663.

⁴⁰⁶ *St George's v S* (n388), 687 (Judge LJ).

⁴⁰⁷ *Re MB* (n 295).

⁴⁰⁸ *Re AA* (n 398).

⁴⁰⁹ *ibid.*, [1] (Mostyn J).

⁴¹⁰ *Norfolk and Norwich Healthcare (NHS) Trust v W* [1997] 1 FCR 269.

⁴¹¹ *ibid.*, 273 (Johnson J).

⁴¹² John Harrington, ‘Time as a Dimension of Medical Law’ [2012] 20 Medical Law Review 491, 511.

⁴¹³ Sara Fovargue and José Miola (n 344), 265.

⁴¹⁴ *Re AA* (n 398); Re P [2013] EWHC 4581.

⁴¹⁵ Sara Fovargue and José Miola (n 399), 254.

⁴¹⁶ Emma Walmsley, ‘Mama Mia! Serious Shortcomings with another ‘(en) forced’ Caesarean Section Case: *Re AA* [2012] EWHC 4378’, [2014] Medical Law Review 1, 6.

of choice for the crudity of compulsion.”⁴¹⁷ It is hard to acquiesce to the prospect that women should have their autonomy subordinated in this way,⁴¹⁸ despite the foetus lacking legal recognition when a mother is not legally required to undergo any medical procedure for her child⁴¹⁹ once it has a legal persona.

V. Conclusion

There are significant inconsistencies in the law regarding intervention in choices during pregnancy and childbirth, and legal maternal responsibilities towards foetuses. The wardship jurisdiction has not been extended to the foetus, civil liability has been precluded, and the Court of Appeal ruled that a woman cannot be guilty of inflicting GBH on her foetus. However, there has been intervention and compulsion in choices during childbirth, and the possibility remains that a woman may be guilty of manslaughter if her foetus dies after birth due to her negligent or unlawful conduct whilst pregnant.

The law should not ‘police pregnancy’ because of the drastic effect intervention has on the autonomy, liberty, and privacy of women, and because it becomes harder to protect the foetus when women are incentivised to seek illegal abortions and are less willing to engage with health and social care. Non-intervention does not signal that harmful conduct in pregnancy is tolerated; instead, it acknowledges that the rights of women are not diminished by pregnancy, and that foetal welfare is not sufficiently improved to justify the infringement of women’s rights. Foetal welfare is better served without intervention pursuing the occasional harmful mother,⁴²⁰ and with education and treatment incentives that are better received by enabling, rather than restricting, choice.⁴²¹

⁴¹⁷ Margaret Brazier (n 289), 375.

⁴¹⁸ Sheila Mclean (n 325), 60.

⁴¹⁹ *ibid.*, 60.

⁴²⁰ Margaret Brazier (n 316), 293.

⁴²¹ Emma Cave (n 343), 37.

Should the law recognise rights in all humans and only in humans?

Aneela Samrat⁴²²

This article examines the widely debated question of whom or what the law should recognise rights in. It rejects the will theory and argues that the law should recognise rights on the basis of the interest theory. The central argument advanced is that the law should ascribe rights to anyone or anything that is capable of having interests, which is determined by the ability to form desires, beliefs, plans, and purposes to which objective goods might be subjectively useful. The distinction between humans and non-humans is unimportant; all humans and animals that are capable of forming interests should have rights. This includes children, the elderly, the mentally ill, the cognitively disabled, and some animals. However, this does not include the brainstem dead, those in permanent vegetative states, some animals, and inanimate objects that are incapable of having interests.

This essay explores who or what the law should recognise as rights-holders. It starts with an overview of the analysis of rights and the will and interest theories. It will then show that the interest theory is the correct conceptual basis for ascribing rights. What an interest is, and who or what is capable of having an interest, is central to determining who or what the law should recognise rights in. What is important is not so much the human versus non-human distinction, but rather the ability to have interests, from which rights stem. It will be argued that most, but perhaps not all, humans and some animals are capable of having interests and hence should have legally recognised rights.

The will and interest theories are regarded as the two primary conceptual theories concerning the nature of rights in the analytical tradition.⁴²³ The will theory interprets rights as linked by the fact that they all protect choices or acts of will.⁴²⁴ It regards a

⁴²² LL.B. Candidate, The University of Manchester, School of Law.

⁴²³ Brian H Bix, *Jurisprudence: Theory and Context* (7th edn Sweet & Maxwell, London 2015) 138.

⁴²⁴ Nigel E Simmonds, *Central Issues in Jurisprudence* (4th edn Sweet & Maxwell, London 2013) 328.

duty only correlative to a claim-right when it is made conditional upon an exercise of will by another.⁴²⁵ Hart equated having a right with the right-holder being given exclusive control over another person's duty.⁴²⁶ This renders the right-holder, to whom the duty is owed, a small-scale sovereign in the area of conduct covered by the duty.⁴²⁷ Under the will theory, rights always confer sovereignty, and so those incapable of exercising sovereignty are incapable of holding rights. This includes children, the cognitively disabled or permanently comatose, and animals. The will theory requires that one be competent of the rational choice to enforce or waive one's right in order to hold a right.⁴²⁸ This denies rights on the basis that potential right-holders do not have the capacity or competence to assert or exercise their rights. Contrastingly, the interest theory extends rights to groups excluded by the will theory: it regards rights as linked by the fact that they serve the right holder's interest.⁴²⁹ Having a right is equated with being the intended beneficiary of another's duty;⁴³⁰ claim-rights are regarded as interests protected by duties.⁴³¹ Rights should be understood in the light of the interest theory.

Something is an interest if its fulfilment enriches an aspect of a right-holder's wellbeing.⁴³² To say that something is in the interest of another is to say that it is good for them in that it will further their purposes and projects.⁴³³ Interests are distinct from the good as they concern only some of what is good for an individual, that which is objectively capable of potentially serving their subjective desires.⁴³⁴ This analysis of interests explains why inanimate objects should not be ascribed interests and hence why

⁴²⁵ Nigel E Simmonds (n 424) 329

⁴²⁶ Brian H Bix (n 423) 138.

⁴²⁷ Brian H Bix (n 423) 138.

⁴²⁸ Mhairi Cowden, Capacity, claims and children's rights (2012) *Contemporary Political Theory* 11(4) 369.

⁴²⁹ Nigel E Simmonds (n 424) 328.

⁴³⁰ Brian H Bix (n 423) 138.

⁴³¹ Nigel E Simmonds (n 424) 329.

⁴³² Rowan Cruft, Mathew Liao and Massimo Renzo, *Philosophical Foundations of Human Rights*, (1st edn Oxford University Press, Oxford 2015) 52.

⁴³³ Robert E Goodin and Diane Gibson, Rights, Young and Old (1997) *Oxford Journal of Legal Studies* 17(2) 189.

⁴³⁴ Robert E Goodin and Diane Gibson (n 433) 190.

the law should not recognise rights in them. To illustrate this, Summer uses the examples of trees, canyons and statues.⁴³⁵ These inanimate objects can be harmed, trees can be felled, canyons filled, statues smashed. It is objectively good for these that destruction be avoided.⁴³⁶ However, these inanimate objects cannot have interests because they are incapable of forming desires, beliefs, plans and purposes, in terms of which the aforementioned objective goods might be subjectively useful.⁴³⁷ Therefore, the debate over what the law should recognise rights in is limited to the living that are capable of forming desires and of planning and feeling, in some capacity, which would render objective goods subjectively useful to them and thus be in their interest.

Having discussed inanimate objects, rights will now be analysed in the context of humans. The debate will then be extended, with the conclusions reached applied to and developed, in respect of animate non-humans. The idea of human rights suggests that all humans have rights. The orthodox view interprets human rights as moral rights belonging to all people simply by virtue of their humanity,⁴³⁸ irrespective of their physical, mental, or social characteristics and development. Generally, a key characteristic of humanity is the ability to hold interests. Regarding rights as contingent on competence, as the will theory does, renders rights “exclusive and exclusionary”.⁴³⁹ For example, the mentally ill are generally excluded from holding rights under this reasoning. Campbell argues that to refuse to speak of rights for the mentally incompetent could give rise to an assumption that their interests are not as important as the interests of right-bearers.⁴⁴⁰ But this cannot be the case in an egalitarian society: human rights are often the most valuable to the vulnerable, who are at higher risk of suffering discrimination.⁴⁴¹ It is these people that are arguably most in need of

⁴³⁵ Robert E Goodin and Diane Gibson (n 433) 190.

⁴³⁶ Robert E Goodin and Diane Gibson (n 433) 190.

⁴³⁷ Robert E Goodin and Diane Gibson (n 433) 190.

⁴³⁸ Rowan Cruft, Mathew Liao and Massimo Renzo (n 432) 46.

⁴³⁹ Michael Freeman, *Lloyd's Introduction to Jurisprudence* (9th edn Sweet & Maxwell, London 2014) 1307.

⁴⁴⁰ Tom Campbell, *The Rights Approach to Mental Illness* (1984) Royal Institute of Philosophy Lecture Series 18 (251).

⁴⁴¹ Michael Freeman (n 439) 1287.

rights to grant them empowerment and the ability to compel accountability.⁴⁴² As such, the law ought to instantiate rights in law in respect of all humans capable of having interests. As Shami Chakrabarti has opined, “a little hard-edged legal protection can make all the difference in the world”.⁴⁴³

Under the interest theory, one can have rights regardless of whether one is capable of exercising or asserting them.⁴⁴⁴ This argument will now be assessed in the context of children’s rights; it will be shown that lacking the capacity and competence to assert and exercise one’s rights should not prevent them from having rights. This is because rights-holders have rights by virtue of their interests. The debate about whether the law should recognise rights in children is “characterised by the divisive concept of capacity”.⁴⁴⁵ The will theory’s approach to human rights is grounded in humans’ capacity for rationally purposive agency.⁴⁴⁶ Proponents O’Neil, Purdy, and Griffin argue that children do not have the capacity to exercise rights,⁴⁴⁷ and as such should not be recognised as having rights. According to Brighthouse, children depart too far from “the liberal model of the rational person”⁴⁴⁸ to justify holding them. That is, the reduced physical and cognitive capacities of children are incorrectly presented as the justification for denying children rights.⁴⁴⁹ Yet interests are regarded as things that “are or will become subjectively useful in the service of possible or probable subjective desires”.⁴⁵⁰ As children have strong, although often transient, desires in the present, they have interests much the same as adults.⁴⁵¹ Therefore, the law should recognise their rights.

Although capacity has been argued not to be relevant to one’s ability to hold a right, it could be submitted that ascribing rights to highly vulnerable people who are unable to exercise them

⁴⁴² Michael Freeman (n 439) 1307.

⁴⁴³ Shami Chakrabarti, *On Liberty*, (Penguin Group, London 2014) 145.

⁴⁴⁴ Michael Freeman (n 439) 1308.

⁴⁴⁵ Mhairi Cowden (n 428) 362.

⁴⁴⁶ Michael Freeman (n 439) 1305.

⁴⁴⁷ Michael Freeman (n 439) 1307.

⁴⁴⁸ Michael Freeman (n 439) 1307.

⁴⁴⁹ Mhairi Cowden (n 428) 364.

⁴⁵⁰ Robert E Goodin and Diane Gibson (n 433) 190.

⁴⁵¹ Robert E Goodin and Diane Gibson (n 433) 191.

is ineffective. If so, the law should not recognise rights in these people. But this conclusion seems absurd, given that it is these vulnerable groups who are most in need of the protection provided by legally recognised rights. Under the interest theory, the fact that people's interests are interpersonally accessible and the claims arising from them are impersonally exercisable⁴⁵² allows for trustees and guardians to be appointed to protect the rights of those incapable of enforcing their own.⁴⁵³ Therefore, the law should recognise rights in all humans capable of having interests regardless of whether the rights-holder can themselves assert or exercise their rights.

Critics such as Guggenheim claim the case for children's rights has no "intellectual foundation",⁴⁵⁴ while Dixon and Nussbaum characterise it as "largely under-theorised".⁴⁵⁵ These criticisms are ironic considering that the denial of children's rights is not grounded on sound philosophical objections, but largely on a concern of, what Griffin feared would be, a dangerous "ballooning"⁴⁵⁶ of the discourse of what the law should recognise as having rights.⁴⁵⁷ This is not a strong enough justification for denying rights to some of the least powerful members of society.

The interest theory's ability to circumvent the issue of capacity in the context of children can be extended to most other so-called problem groups. However, some mistakenly argue that rather than providing a crucial test case, the example of children is merely a quirk.⁴⁵⁸ This view is based on a version of the will theory where by distinguishing between capacity and competence, children can hold rights. On the other hand, others such as the mentally infirm, elderly, or the cognitively disabled, cannot. Cowden illustrates this distinction with the example that: both a student and a turtle may currently lack the competence to speak Russian, but only the student has the capacity to one day become competent in speaking Russian.⁴⁵⁹ Similarly, it is assumed that children will develop the

⁴⁵² Robert E Goodin and Diane Gibson (n 433) 192.

⁴⁵³ Robert E Goodin and Diane Gibson (n 433) 193.

⁴⁵⁴ Michael Freeman (n 439) 1307.

⁴⁵⁵ Michael Freeman (n 439) 1307.

⁴⁵⁶ Michael Freeman (n 439) 1307.

⁴⁵⁷ Michael Freeman (n 439) 1307.

⁴⁵⁸ Robert E Goodin and Diane Gibson (n 439) 195.

⁴⁵⁹ Mhairi Cowden (n 428) 366.

competencies to make moral decisions in the future. The capacity of the mentally infirm, elderly, and the cognitively disabled, however, appears to be permanently impaired; under this variation of the will theory, they are denied rights. But under the interest theory, this argument is wholly irrelevant. Neither competence nor capacity determines a right – interests do. Admittedly, it may be more difficult for guardians to ascertain the interests of people who lose capacity later in life than those of children because the former group have more focused subjective desires.⁴⁶⁰ But this is a difficulty for, not with, these groups holding rights.

Nevertheless, the requirement that interest-holders be able to form desires, beliefs, plans, and purposes, does pose a problem for the brainstem dead and those in permanent vegetative states. A person who is brainstem dead no longer has any activity in their brainstem; they have permanently lost the potential for consciousness and have no chance of recovery.⁴⁶¹ Others may owe duties to them, but the law should not recognise rights in respect of them, as they are incapable of having interests for rights to serve. This might appear to conflict with the human rights argument that all humans have rights by virtue of their humanity, but being human is arguably based on having at least some cognitive function that the brainstem dead no longer possess.

A person in a permanently vegetative state is awake but shows no signs of awareness; they are incapable of showing any meaningful responses or signs of experiencing emotions.⁴⁶² So like the brainstem dead, they are incapable of having interests. This view is supported by the words of Lord Mustill in *Airedale NHS Trust v Bland*.⁴⁶³ The case concerned Anthony Bland, who was reduced to a persistent vegetative state and subsequently kept alive on life support machines for three years. With the consent of Bland's parents, the hospital applied for a declaration to lawfully discontinue all life-sustaining treatment. Based on the best interests test, Lord

⁴⁶⁰ Robert E Goodin and Diane Gibson (n 433) 194.

⁴⁶¹ <<http://www.nhs.uk/Conditions/Brain-death/Pages/Introduction.aspx>> accessed 18 January 2016.

⁴⁶² <<http://www.nhs.uk/conditions/vegetative-state/Pages/Introduction.aspx>> accessed 18 January 2016.

⁴⁶³ *Airedale NHS Trust v Bland* [1993] AC 789.

Mustill famously stated that “the distressing truth which must not be shirked is that the proposed conduct is not in the best interests of Anthony Bland, for he had no best interests of any kind”.⁴⁶⁴ As such, the permanently vegetative are incapable of holding interests and therefore, under the interest theory, rights.

There are clearly right and wrong ways for the brainstem dead and permanently vegetative to be treated,⁴⁶⁵ but what is objectively good cannot be subjectively tailored to them, as they are devoid of distinct cognitive human character. We may have duties to the brainstem dead and permanently vegetative, but they are incapable of having interests, so the law should not recognise rights in them.

Much of what has already been discussed in respect of hard cases is also applicable to animate non-humans. The lack of capacity of animals to assert and exercise their rights does not exclude them from having rights, just as it does not prevent children or the mentally ill from having rights. Animals with the cognitive ability to form desires, beliefs, plans, and purposes, will have interests in relation to which objective goods can be subjectively applied. Under the interest theory, they should, therefore, have rights.

But Frey argues that what is believed is that a given sentence is true and as animals lack linguistic proficiency they cannot believe that any sentence is true. Firstly, in response to this, not all animals lack linguistic proficiency. Some animals such as chimpanzees have a degree of linguistic competence. Notwithstanding this, it is not necessarily the case that beliefs only exist in those that can articulate them in language. The cognitively disabled and infants often also lack linguistic proficiency but have been proven to have interests. Thus, so should animals that are capable of holding beliefs and desires. Moreover, Wittgenstein notes that “if a lion could speak, we could not understand him”.⁴⁶⁶ That is, just because animals process beliefs and desires in ways that humans cannot comprehend linguistically, does not mean animals lack beliefs and desires and hence interests worthy of the protection of legally recognised rights.

⁴⁶⁴ *Airedale NHS Trust v Bland* [1993] AC 789 at 46.

⁴⁶⁵ Robert E Goodin and Diane Gibson (n 433) 196.

⁴⁶⁶ Ludwig Wittgenstein, Gertrude E M Anscombe, Peter M S Hacker and Joachim Schulte, *Philosophical Investigations*, (4th edn Blackwell Publishing Ltd, West Sussex 2009) 223.

It can be argued that if we have no access to, or understanding of, the beliefs and desires of animals, then it is unclear on what basis animals have beliefs and hence rights. However, denying animals legally recognised rights on the basis that we cannot say what animals believe, is akin to arguing that it is difficult to represent the interests of older people who lose cognitive functioning.⁴⁶⁷ In other words, this is a problem for, and not with, the idea of animals having rights. It is reasonable to assume that at least some non-human animals have interests. Where that assumption is reasonable, rights should be attributed to these non-human animals on the basis of the interest theory. It should be stressed that animals capable of holding interests should have legally recognised rights.

Few deny that beliefs and desires should be attributed to animals, particularly mammalian animals,⁴⁶⁸ as there are strong evolutionary links and behavioural similarities between them and humans.⁴⁶⁹ Feinberg opines that many of these higher animals have interests because they have “appetites, conative urges, and rudimentary purposes, the integrated satisfaction of which constitutes their welfare or good”.⁴⁷⁰ Indeed, evolutionary theory indicates that mammalian animals often behave as they do because of their desires and beliefs. Moreover, Stich remarks that “it is hard to believe that belief-desire psychology could explain human behaviour, but not animal behaviour”,⁴⁷¹ and subsequently concludes that “if humans have beliefs, so too do animals”.⁴⁷² If animals have beliefs and desires that constitute interests, then under the interest theory, they should have rights. Regan argues that the burden of proof falls on those that deny that animals have beliefs and desires and hence interests that afford them rights.⁴⁷³ But to deny that animals have rights, one must show that “it is reasonable to deny

⁴⁶⁷ Tom Regan, *The Case for Animal Rights*, (2nd edn University of California Press, Berkeley 2004) 37.

⁴⁶⁸ Tom Regan (n 467) 35.

⁴⁶⁹ Tom Regan (n 467) 36.

⁴⁷⁰ Joel Feinberg, “The Rights of Animals and Unborn Generations” in Ruff Shafer-Landau (ed), *Ethical Theory: An Anthology* (2nd edn Wiley-Blackwell, West Sussex 2013).

⁴⁷¹ Tom Regan (n 467) 36.

⁴⁷² Tom Regan (n 467) 36.

⁴⁷³ Tom Regan (n 467) 37.

that the belief-desire theory applies to animal behaviour without implying that the same is true in the case of human behaviour.”⁴⁷⁴

Animals that can form beliefs and desires have interests and therefore rights. Deciding exactly which species of animal have the ability to form desires, beliefs, plans, and purposes, should be left to experts in the cognitive abilities of animals. It is not so important to this essay as to which animals can be rights-holders, but more so that some animals do indeed have rights, and therefore should be afforded legally recognised rights, on the same basis as humans. It should also be noted that saying that the law should recognise rights in some animals does not mean that these animals are deserving of all the rights that adult humans have. The conclusion is that some animals should have legally recognised rights on account of their interests, but the scope of these rights is beyond the remit of this inquiry.

What is important is not the human versus non-human distinction, but rather the ability to have interests, from which rights stem. As explained, the law should recognise rights in all interest-holders, whether human or not. This includes most, but not all, humans and some animals; specifically, it covers those capable of forming desires, beliefs, plans, and purposes, in terms of which objective goods might be subjectively useful and thus be in their interests.

⁴⁷⁴ Tom Regan (n 467) 37.

An exploration into the factors shaping victim reporting of partner abuse to the police

Victoria Smith⁴⁷⁵

Victims of partner abuse are indicated to be amongst the least likely to report their victimisation to police. Using Scottish Crime and Justice Survey data, this report explores victim reporting of partner abuse to the police, presenting bivariate and logistic regression analyses of the factors involved. Findings indicate that female victims, victims whose children were present or involved in the abuse, victims who felt victimised, and victims experiencing multiple physical effects are the most likely to report. This has implications for determining valid victimisation rates, identifying and apprehending offenders, and ensuring victims are aware of, and receive, support and services.

I. Introduction and literature review

Although differences in the approach of measuring partner abuse⁴⁷⁶ and recurring under reporting⁴⁷⁷ make gaining accurate measurements problematic, latest figures show that from 2014 to 2015 there were 59,882 incidents of domestic abuse recorded by the police in Scotland. This is an increase of 2.5% from 2013 to 2014, suggesting this remains a significant problem.⁴⁷⁸ Additionally, from 2014 to 2015 in England and Wales, the figure was 943,628; however, unlike the Scottish figure, this includes abuse by family members.⁴⁷⁹ Official crime rates are generally used to determine the response received by victims within the criminal justice system; nevertheless, self-report surveys of victimisation indicate that many do not report their victimisation to law enforcement agencies.⁴⁸⁰

⁴⁷⁵ LL.M. Candidate, The University of Manchester, School of Law.

⁴⁷⁶ Sandra Walklate, *Gender, Crime and Criminal Justice* (2nd edition, Willan 2004).

⁴⁷⁷ Nicola Harwin, 'Putting a Stop to Domestic Violence in the United Kingdom: Challenges and Opportunities' (2006) 12 (6) 'Violence Against Women' 556.

⁴⁷⁸ Scottish Government, 'Domestic Abuse Recorded by the Police in Scotland 2013-14 & 2014-15' (2015) <<http://www.gov.scot/Resource/0048/00487981.pdf>> accessed 24 March 2016.

⁴⁷⁹ John Woodhouse & Noel Dempsey, *Domestic violence in England and Wales* HoC Briefing Paper 6337 (2016) <<http://researchbriefings.files.parliament.uk/documents/SN06337/SN06337.pdf>> accessed 08 May 2016.

⁴⁸⁰ Heather Zaykowski, 'Reconceptualizing Victimization and Victimization Responses', (2015) 61 (2) 'Crime & Delinquency' 271, 271; Jennifer Truman & Michael Planty, *Criminal*

This ‘dark figure’ of crime serves to demonstrate victims’ reluctance to report their victimisation to the police.⁴⁸¹

Scottish national policy and legislation over the last decade has criminalised abusive behaviour and introduced several civil measures aiming to address partner abuse. The Scottish Government’s Violence Against Women team have also developed key frameworks to aid awareness among practitioners who may encounter female victims of partner abuse. However, victims of partner abuse are suggested as amongst the least likely to report their victimisation to the police.⁴⁸² For example, the Scottish Crime and Justice Survey (SCJS) 2012/13 wave revealed that only around 13% of respondents experiencing partner abuse within the last 12 months told the police about their most recent incident,⁴⁸³ compared to around 48% for violent crimes.⁴⁸⁴ This highlights the lack of reporting surrounding partner abuse and renders this report particularly helpful in understanding the research in this area.

Citizens may be perceived as the ‘gatekeepers’ in controlling crime, given that reporting is viewed to be the initial stage in the criminal justice process.⁴⁸⁵ Research indicates that it is more probable that victims will utilise informal networks such as friends or family, than formal sources of support, including the police.⁴⁸⁶ This disparity is significant since it impacts upon “determining valid rates of victimization, identifying and apprehending offenders, and ensuring that crime victims are made

Victimization, 2011 US DOJ Washington DC Bureau of Justice Statistics, NCJ 239437 (2012) <<http://www.bjs.gov/content/pub/pdf/cv11.pdf>> accessed 07 April 2016.

⁴⁸¹ Wesley Skogan, ‘Reporting crimes to the police: The status of world research’ (1984) 21 ‘Journal of Research in Crime and Delinquency’ 113.

⁴⁸² Sarah MacQueen & Paul Norris, ‘Police awareness and involvement in cases of domestic and partner abuse’ (2016) 26 (1) ‘Policing and Society’ 55, 55.

⁴⁸³ Sarah MacQueen, ‘2012/13 Scottish Crime and Justice Survey: Partner Abuse’, Scottish Government Social Research (2014) <<http://www.gov.scot/Resource/0045/00455441.pdf>> accessed 25 March 2016.

⁴⁸⁴ Scottish Government, ‘Scottish Crime and Justice Survey 2012/13: Main Findings’ Scottish Government (2014) <<http://www.gov.scot/Resource/0044/00447271.pdf>> accessed 25 March 2016.

⁴⁸⁵ Zaykowski (n 480); Michael Gottfredson & Don Gottfredson, *Decision making in criminal justice* (2nd edition, Plenum 1998).

⁴⁸⁶ Ann Coker, Christina Derrick, Julia Lumpkin, Timothy Aldrich & Robert Oldendick, ‘Help-seeking for intimate partner violence and forced sex in South Carolina’ (2000) 19 ‘American Journal of Preventive Medicine’ 316; Sarah Ullman, ‘Mental health services seeking in sexual assault victims’ (2007) 30 ‘Women & Therapy’ 61.

aware of and receive needed support and services.”⁴⁸⁷ Indeed, the European Commission⁴⁸⁸ prioritised encouraging victims and witnesses to report violence against women and/or children to the relevant authorities and institutions. Understanding the factors which shape victim reporting of partner abuse to the police is therefore pertinent, particularly considering the low police visibility due to such crimes’ private nature.⁴⁸⁹ It is also suggested that the widespread “social, personal, and economic costs of crime and violence, as well as the lasting health effects” make understanding victims’ responses and the types of help sought, key to addressing partner abuse.⁴⁹⁰

In this report, ‘partner abuse’ is consistent with the SCJS questionnaire approach. The definition used in the questionnaire is similar to that in the ‘Joint Protocol between Police Scotland and Crown Office and Procurator Fiscal Service’⁴⁹¹ which is “any form of physical, sexual or mental and emotional abuse which might amount to criminal conduct and which takes place within the context of a relationship. The relationship will be between partners (married, cohabiting, civil partnership or otherwise) or ex-partners. The abuse can be committed in the home or elsewhere.” This definition acknowledges male violence towards women, female violence towards men, and violence between partners or ex-partners in same-sex relationships.

There exists a well-established and expanding literature that examines factors influencing victim reporting to the police. It indicates that this decision is influenced by numerous factors,

⁴⁸⁷ Bradford Reynolds & Christine Englebrecht, ‘Informal and formal help-seeking decisions of stalking victims in the United States’ (2014) 41 (10) ‘Criminal Justice and Behaviour’ 1178, 1178-1179.

⁴⁸⁸ European Commission., *Istanbul Convention: combatting violence against women*. European Commission (2016) <http://ec.europa.eu/justice/newsroom/files/gender-equality/factsheet_istanbul_convention_web_en.pdf> accessed 8 May 2016.

⁴⁸⁹ Donald Black, ‘The mobilisation of law’ (1973) 2 (1) ‘The Journal of Legal Studies’ 125.

⁴⁹⁰ Catherine Kaukinen, Silke Meyer & Caroline Akers, ‘Status Compatibility and Help-Seeking Behaviours Among Female Intimate Partner Violence Victims’ (2013) 28 (3) ‘Journal of Interpersonal Violence’ 577, 577.

⁴⁹¹ Crown Office and Procurator Fiscal Service, *Joint Protocol between Police Scotland and Crown Office and Procurator Fiscal Service* COPFS (2013)

<http://www.copfs.gov.uk/images/Documents/Prosecution_Policy_Guidance/Protocols_and_Memorandum_of_Understanding/Joint%20Domestic%20Abuse%20Protocol%20-%20Oct%2013.pdf> accessed 18 April 2016.

including certain characteristics of the relationship, perceptions and confidence in the criminal justice system (CJS) and the police, and specific victim characteristics. Nonetheless, despite a recent proliferation in analyses of the reporting of partner abuse to the police, and a heightened focus on the types of help that female victims seek, it arguably remains under-researched.⁴⁹²

(i) Characteristics of the relationship

The relationship between the victim and the offender is vital in influencing the victim's decision to report.⁴⁹³ It is suggested that "victims may perceive more benefits in reporting crime when they know the offender if they believe that there is greater chance of reprisal, or that the police are more likely to catch and punish the known offender."⁴⁹⁴ Nevertheless, Black⁴⁹⁵ suggests that a broad structural relationship exists between 'relational distance' and the activation of law; the closer the relationship between victim and offender, the less likely the victim will report. This theory may provide an explanation as to why victims of partner abuse are less likely to report. Consistent with this, the likelihood of reporting is found to increase when the offender is a stranger, particularly regarding sexual assault.⁴⁹⁶ However, the exact influence on reporting is debated, and appears to depend on the type of victimisation.⁴⁹⁷ It, therefore, appears that further research is necessary to clarify this position.

Victims are found to be more likely to forgive rather than admit to the criminal behaviour of an intimate partner. Often, crimes involving known offenders, such as partner abuse, are considered to be private with resolution being sought through informal social

⁴⁹² MacQueen & Norris (n 482) 56-57.

⁴⁹³ Gottfredson & Gottfredson (n 485).; Martin Greenberg & Barry Ruback, *After the Crime: Victim Decision Making* (Plenum 1992).

⁴⁹⁴ Elizabeth Estienne & Melissa Morabito, 'Understanding differences in crime reporting practices: a comparative approach' (2016) 40 (2) 'International Journal of Comparative and Applied Criminal Justice' 123, 125.

⁴⁹⁵ Donald Black, *The behaviour of law* (Academic Press 1976).

⁴⁹⁶ Bonnie Fisher, Leah Daigle, Francis Cullen & Michael Turner, 'Reporting Sexual Victimisation To The Police And Others' (2003) 30 (1) 'Criminal Justice and Behaviour' 6.

⁴⁹⁷ Reynolds & Englebrecht (n 487), 1181.

control.⁴⁹⁸ Consequently, the police are less likely to be called. Additionally, research in the area of partner abuse highlights that marriage and cohabitation decrease the likelihood of reporting.⁴⁹⁹ These are factors which may be used to indicate the ‘relational distance’ between the victim and their partner.

Unique to partner abuse, analyses of the presence of children and their witnessing of abuse is established as positively influencing the likelihood of police reporting.⁵⁰⁰ For example, Chang and colleagues⁵⁰¹ highlight that threats to the safety or well-being of a victim’s children drives the victim to act in a way which will protect them. Conversely, the involvement of children may not have this effect if it is believed that there may be detrimental consequences to the victim and their children.⁵⁰²

⁴⁹⁸ Richard Felson, Steven Messner, Antony Hoskin, & Glenn Deane, ‘Reasons for Reporting and Not Reporting Domestic Violence to the Police’ (2002) 40 (3) ‘Criminology’ 617; Roger Tarling, & Katie Morris, ‘Reporting Crime to the Police’ (2010) 50 (3) ‘British Journal of Criminology’ 474.

⁴⁹⁹ Catriona Mirrlees-Black, *Domestic violence: findings from a new British crime survey self-completion questionnaire*. Home Office Research Study No. 191. London: Home Office (1999); Caroline Akers, & Catherine Kaukinen, ‘The police reporting behavior of intimate partner violence victims’ (2009) ‘Journal of Family Violence’ 24 159171.

⁵⁰⁰ Mirrlees-Black (n 499); Akers & Kaukinen (n 499); MacQueen & Norris (n 482).

⁵⁰¹ Judy Chang, Diane Dado, Lynn Hawker, Patricia Cluss, Raquel Buranosky, Leslie Slagel, Melissa McNeil & Sarah Scholle, ‘Understanding turning points in intimate partner violence: factors and circumstances leading women victims towards change’ (2010) 19 (2) ‘Journal of Women’s Health’ 251259.

⁵⁰² Therese Zink, Nancy Elder & Jeff Jacobson, ‘How children affect the mother/victim’s process in intimate partner violence’ (2003) 157 (6) ‘Archives of pediatrics & adolescent medicine’ 587.

(ii) Specific attributes related to the incident

The rational choice model is the most commonly proposed approach when considering a victim's decision to report crime to the police.⁵⁰³ Accordingly, crime reporting is underpinned by weighing up the costs and benefits of this action,⁵⁰⁴ in which offence seriousness generally plays a crucial role.⁵⁰⁵ This is consistently regarded as one of the most significant predictors and is thought to be determined by either the presence and extent of injuries sustained, or the victim's subjective view of how serious the offence was. Victims of domestic violence have been found more likely to report their experiences to the police when injuries were significant,⁵⁰⁶ whilst the use of a weapon is indicated to increase the likelihood of reporting to the police.⁵⁰⁷

Furthermore, significant injuries are also more likely to lead to the victim to label their experience as victimisation. This is recognised as an important predictor of victim reporting; for example, in cases of sexual assault, "victims are less likely to label their experience as criminal if their victimization did not fit the stereotypical account of an assault."⁵⁰⁸ Regarding partner abuse, victims' perceptions of their experiences influence their inclination to take up services;⁵⁰⁹ it is therefore argued that the process of victimisation and crime reporting is underpinned by the individual

⁵⁰³ Estienne & Morabito (n 494) 124.

⁵⁰⁴ Block, R., 'Why Notify the Police The victim's decision to notify the police of an assault' (1974) 11 (4) 'Criminology' 555; Gottfredson & Gottfredson (n 485).

⁵⁰⁵ Wesley Skogan, 'Citizen Reporting of Crime: Some National Panel Data' (1976) 13 (4) 'Criminology' 535.

⁵⁰⁶ Ronet Bachman, & Ann Coker, 'Police involvement in domestic violence: The interactive effects of victim injury, offender's history of violence, and race' (1995) 10 'Violence and Victims' 91; Heike Goudriaan, James Lynch & Paul Nieuwbeerta, 'Reporting to the police in western nations: a theoretical analysis of the effects of social context' (2004) 21 (4) 'Justice Quarterly' 933.

⁵⁰⁷ Kim Davies, Carolyn Block, & Jacquelyn Campbell, 'Seeking help from the police: Battered women's decisions and experiences' (2007) 20 'Criminal Justice Studies' 15.

⁵⁰⁸ Reyns & Englebrecht (n 487), 1180; Jody Clay-Warner, & Jennifer McMahon-Howard, 'Rape Reporting: "Classic Rape" and the Behavior of Law' (2009) 24 (6) 'Violence and Victims' 723; Karen Weiss, "'Boys Will Be Boys" and Other Gendered Accounts: An Exploration of Victims' Excuses and Justifications for Unwanted Sexual Contact and Coercion' (2009) 15 (7) 'Violence Against Women' 810.

⁵⁰⁹ Mirrlees-Black (n 499).

classing themselves as a victim as opposed to claiming victimisation.⁵¹⁰

(iii) Perceptions and confidence in the criminal justice system and the police

Despite the importance of incident-level characteristics, other factors such as victim characteristics, and factors relating to the police may also shape reporting.⁵¹¹ Trust in the police has been defined as the belief that the police are competent in their role and that the interest of the public is central to what they do.⁵¹² Evidence suggests that perceptions of the police surrounding their capabilities, fairness, and effectiveness influence crime reporting.⁵¹³ In addition, trust in justice generally has been found to be an important determinant of whether an individual cooperates with the criminal justice system, including whether they report crime to the police.⁵¹⁴

⁵¹⁰ William Felstiner, Richard Abel & Austin Sarat, 'The Emergence and Transformation of Disputes: Naming, Blaming, Claiming...' (1980/1981) 15 (3-4) 'Law and Society Review' 631.

⁵¹¹ Estienne & Morabito (n 494).

⁵¹² Jonathon Jackson, Ben Bradford, Betsy Stanko & Katrin Hohl, *Just Authority? Trust In The Police In England and Wales* (Routledge, 2013).

⁵¹³ Ben Bradford & Jonathon Jackson, 'Cooperating with police: Social control and the reproduction of police legitimacy', *London School of Economics working paper* (2010) <<http://ssrn.com/abstract=1640958>> accessed 04 April 2016; Ziggy Macdonald, 'Revisiting the dark figure: a microeconomic analysis of the under-reporting of property crime and its implications' (2001) 41 (1) 'British Journal of Criminology' 127.

⁵¹⁴ European Social Survey, 'Policing by Consent: Understanding the Dynamics of Police Power and Legitimacy', ESS Country Specific Topline Results Series Issue 1 (UK) (2012) <<http://ssrn.com/abstract=2168702>> accessed 01 March 2016.; Jonathon Jackson, Ben Bradford, Mike Hough, Andy Myhill, Paul Quinton & Tom Tyler, 'Why do people comply with the law? Legitimacy and the Influence of Legal Institutions' (2012) 52 (6) 'British Journal of Criminology' 1051.

(iv) Victim characteristics

Victims' characteristics also influence whether they report crime to the police. For example, women are more likely than men to report crimes.⁵¹⁵ This may be due to the traditional view that they are "more fragile, and thus, are expected to seek outside help for protection."⁵¹⁶ Additionally, there exists the view that men should cope with victimisation by themselves.⁵¹⁷ Ethnicity has also been examined, with ethnic minority victims being found less likely to report crimes than non-ethnic minorities.⁵¹⁸ Findings differ between the UK and US; analysis indicates that ethnic minorities are less likely⁵¹⁹ to report crimes in the UK compared with the US.⁵²⁰ However, difficulty in separating this finding from socio-economic disadvantage has been acknowledged.⁵²¹

The socio-economic status of victims also shapes reporting, though its impact is less clear, potentially due to different measures of income.⁵²² Some studies have found that victims with higher socio-economic status, measured according to social classification, household income and employment status, are more likely to report personal and property crime.⁵²³ Conversely, higher status victims may feel embarrassed by their involvement in the CJS,⁵²⁴ making them less likely to report. In relation to partner abuse, analyses in

⁵¹⁵ Catherine Kaukinen, 'The help-seeking decisions of violent crime victims: An examination of the direct and conditional effects of gender and the victim-offender relationship' (2002) 17 *'Journal of Interpersonal Violence'* 432.; Goudriaan, Lynch & Nieuwbeerta (n 506).

⁵¹⁶ Estienne & Morabito (n 494), 125.

⁵¹⁷ Zavala, E., 'Deviant lifestyles and the reporting of physical victimization to the police' (2010) 25 (1) *'Journal of Family Violence'* 23.

⁵¹⁸ Heike Goudriaan, Karin Wittebrood & Paul Nieuwbeerta, 'Neighbourhood characteristics and reporting crime: effects of social cohesion, confidence in police effectiveness and socio-economic disadvantage' (2006) 46 (4) *'British Journal of Criminology'* 719.; Wesley Skogan, *'Contacts between police and public: findings from the 1992 British Crime Survey'* Home Office Research Study No. 13. London: Home Office 1994.

⁵¹⁹ Aisha Gill, 'Voicing the silent fear: South Asian women's experiences of domestic violence' (2004) 43 (5) *'The Howard Journal of Criminal Justice'* 465.

⁵²⁰ Bachman & Coker (n 506).

⁵²¹ Block (n 504).

⁵²² Estienne & Morabito (n 494).

⁵²³ Skogan (n 518).; Goudriaan, Lynch & Nieuwbeerta (n 506).

⁵²⁴ Akers & Kaukinen (n 499).

England and Wales indicate that victims with poor financial circumstances are more likely to report.⁵²⁵

The likelihood of reporting increases with age;⁵²⁶ this has also been established in studies relating to partner abuse.⁵²⁷ On the other hand, Block⁵²⁸ and Skogan⁵²⁹ suggest a non-linear effect of age on police reporting, as reflected in partner abuse literature.⁵³⁰ A more recent study demonstrates a curvilinear relationship where likelihood of reporting increases up until a critical age and then begins to decrease.⁵³¹

Moreover, the community is important in understanding victim reporting. Previous analyses have shown that victims living in the most affluent and most disadvantaged neighbourhoods have the lowest rates of reporting assault to the police.⁵³² Additionally, higher levels of social cohesion within communities have been found to slightly increase the probability of victim reporting, whereas increased levels of socio-economic disadvantage have led to decreased probability.⁵³³

Recent analyses undertaken by MacQueen and Norris⁵³⁴ of the SCJS found that female victims, victims without employment, victims experiencing multiple abuse, and victims whose children witness abuse, were most likely to come to the attention of the police. However, their study focused more broadly on police awareness of partner abuse than victim reporting on its own. Additionally, analysis was limited to the 2008/2009 wave, resulting in a small number of cases. This report focuses specifically on victim reporting and includes both males and females. It also examines perceptions of victimisation, which is identified as potentially important in shaping reporting but was excluded from

⁵²⁵ Mirrlees-Black (n 499).

⁵²⁶ Skogan (n 518); MacDonald (n 513); Goudriaan, Lynch & Nieuwebeerta (n 506).

⁵²⁷ Mirrlees-Black (n 499); Tarling & Morris (n 498).

⁵²⁸ Block (n 504).

⁵²⁹ Skogan (n 505).

⁵³⁰ Richard Felson & Paul-Philippe Paré 'The reporting of domestic violence and sexual assault by nonstrangers to the police' (2005) 67 'Journal of Marriage and Family' 597.

⁵³¹ Akers & Kaukinen (n 499).

⁵³² Eric Baumer, 'Neighbourhood disadvantage and police notification by victims of violence' (2002) 40 'Criminology' 579.

⁵³³ Goudriaan, Wittebrood & Nieuwebeerta (n 518).

⁵³⁴ MacQueen & Norris (n 482).

their analyses. Factors surrounding children have also been analysed in more detail: the report distinguishes between a child's presence and involvement rather than simply children seeing or hearing abuse.

II. Data and methods

(i) Dataset

This report analyses merged data from three waves of the SCJS gathered between 2008 and 2011. Until 2011, domestic abuse data was deposited at the UK Data Service subject to standard conditions. It is a survey of public experiences and perceptions of crime in Scotland carried out annually, predominantly using stratified random sampling, although sampling is clustered in more rural areas. It involves interviews with around 16,000 adults aged 16 or over living at private residential addresses in Scotland.⁵³⁵ However, the majority of the sample each year was excluded from analysis since respondents had either not been a victim of partner abuse or because their abuse did not take place within the survey reference period. Although it is a large survey, only 321 people on average identified as victims each year, with a minority reporting their abuse to the police. Therefore, the three datasets had to be combined to provide sufficient statistical power to be able to conduct multivariate analyses. The final dataset available for analysis consisted of 964 cases.

The SCJS sampling frame, similar to other national surveys, excludes communal establishments and temporary accommodation;⁵³⁶ this potentially excludes victims of serious domestic violence in emergency temporary accommodation.⁵³⁷ Nevertheless, it is useful for examining partner abuse, as it adequately addresses issues levelled at using surveys in this way; for

⁵³⁵ Page, L. & Twist, N. 2010/11 *Scottish Crime and Justice Survey: Technical Report*. Scottish Government Social Research (2011) http://doc.ukdataservice.ac.uk/doc/7229/mrdoc/pdf/7229_scjs_s4_2010-11_technical_report.pdf accessed 25 March 2016.

⁵³⁶ *ibid.*

⁵³⁷ Sylvia Walby & Andrew Myhill, 'New Survey Methodologies in Researching Violence Against Women' (2001) 41 (3) 'British Journal of Criminology' 502.

example, issues of sensitivity⁵³⁸ are addressed by including the questions surrounding partner abuse within the self-completion questionnaire, thereby ensuring privacy for the respondents.⁵³⁹ Additionally, to deal with issues surrounding perceptions of victimisation, the use of the term ‘partner abuse’ was excluded before asking questions about their experiences.

(ii) Dependent variable

To ascertain if the victim reported their abuse to the police, a new dependent variable was created. As highlighted in the literature this is important for numerous reasons, including aiding the services available to victims, increasing reporting rates, and allowing assessment of police resources. Victims were asked ‘How did the Police come to know about the most recent incident?’; if they answered, ‘from them’, then the new variable was coded ‘1’ and if not from them, then it was coded ‘0’. Victims were also asked ‘Did the Police come to know about the most recent incident when your partner or ex-partner did these things to you?’; if answering no, those responses were also coded ‘0’. Only 18% of victims (n = 174) responded yes.

(iii) Explanatory variables

Existing literature identifies several variables potentially influential in shaping victim reporting to the police, which were included as independent variables. Gender, perceptions of victimisation, and living with a partner at the time of the most recent incident, are all nominal dichotomous variables, and age is split into bands.

Two variables relating to socio-economic status were included: household income and the National Statistics Socio-economic Classification. However, income exhibited high levels of non-response. Income was recoded into a binary variable ‘Below average income’ including any income up to £19,000 and ‘Above

⁵³⁸ Siri Thoresen & Carolina Overlien, ‘Trauma victim: yes or no?: why it may be difficult to answer questions regarding violence, sexual abuse and other traumatic events’ (2009) 15 (6) ‘Violence Against Women’ 699.

⁵³⁹ MacQueen & Norris (n 482), 62.

average income' including any income over £20,000. This was based on the UK median income for 2010-2011 of around £20,000.⁵⁴⁰ For the National Statistics Socio-economic Classification, a broad six occupational category version was used which included: managerial and professional; intermediate; supervisory and technical; routine and manual; never worked and long-term unemployed, and full-time student.

A family environment variable was created to assess the presence of children in the household and whether they had been exposed to the most recent incident of abuse. This variable took four values: households without children; households where children were living; households where children were present at the most recent incident; and households containing children where the children were involved in the most recent incident, which includes seeing, hearing, being involved, or being hurt or injured during the incident.

To examine the severity of the incident, two summative measures were created to indicate the number of physical and psychological effects experienced by the victim because of their most recent experience of abuse.

To explore attitudes towards the CJS, three questions were examined: 'How confident are you that the Scottish Criminal Justice System as a whole: is effective in bringing people who commit crimes to justice; deals with cases promptly and efficiently; provides a good standard of service for victims of crime?' Similarly, to assess attitudes towards the local police, three questions were examined: 'How confident are you in your local police force's ability to: respond quickly to appropriate calls and information from the public; deal with incidents as they occur; investigate incidents after they occur?' For these questions, respondents could answer 'Very confident', 'Fairly confident', 'Not very confident', or 'Not at all confident'.

Finally, the literature indicated that neighbourhood socio-economic disadvantage may be a crucial factor to explore. As such, the Scottish Index of Multiple Deprivation quintiles, which are

⁵⁴⁰ HMRC, (2013) *Income and tax by county and region, 2010-11*. ONS <<https://www.gov.uk/government/statistics/income-and-tax-by-county-and-region-2010-to-2011>> accessed 08 April 2016.

included in the SCJS datasets, were utilised to examine the importance of this factor.

(iv) *Analytical strategy*

The assessment of factors involved two stages of analysis, given the predominantly exploratory nature of the report. Bivariate analysis is presented in Tables 1-4, focusing on explanatory variables selected for theoretical reasons based on the existing literature; these illustrate the proportion of victims who reported their most recent incident of abuse to the police across the various explanatory variables. They also indicate statistical significance using chi-square as a goodness-of-fit test,⁵⁴¹ in which a significant chi-square, below .05, suggests that the observed difference is unlikely to have occurred by chance. T-tests were implemented to test the significance of the number of physical and psychological effects.⁵⁴²

The analysis of these variables informed the second stage, whereby only those meeting a threshold of bivariate association were used as part of a logistic regression model, presented in Table 5. Given the dichotomous dependent variable, this was implemented to control for potential links between the explanatory variables and to ascertain which factors have the most prominent associations with reporting when considering these variables together. This two-staged analysis was proceeded with because, although three waves were merged, this still only provided a small number of victims who reported the most recent incident of partner abuse to the police ($n = 174$), which is problematic for the fit of logistic regression. Although this strategy may overfit to the observed sample, the sparsity of data and the exploratory nature of this analysis justify this approach.

According to Agresti,⁵⁴³ there should be no more than one predictor for every ten cases that present the condition being predicted. However, some suggest the possibility of a more stringent

⁵⁴¹ David Weisburd & Chester Britt, *Statistics in Criminal Justice* (Springer 2014).

⁵⁴² Alan Agresti & Barbara Finlay, *Statistical Methods for the Social Sciences* (4th edition, Pearson 2014).

⁵⁴³ Alan Agresti, *Analysis of Ordinal Categorical Data*. (2nd edition Wiley 2010).

20 cases per predictor.⁵⁴⁴ The final model presents odds ratios to support interpretation, whereby an odds ratio value above one indicates increased likelihood of reporting to the police, whereas a value below one signifies decreased likelihood of reporting.⁵⁴⁵ The model predicts a positive outcome of the respondent reporting the most recent incident to the police.

Survey weights were employed for all analyses, given the complex design of the SCJS. This was done to adjust for its sample design⁵⁴⁶ and to compensate for factors which may have made the sample data unrepresentative of the population. If they were not employed, the estimates of population characteristics in this analysis may have been inaccurate or biased.⁵⁴⁷

Although income has been highlighted as relevant⁵⁴⁸ and is significant at the bivariate level – demonstrating that victims with below average income are more likely to report to the police – a regression model which included income was not significant and did not fundamentally change the significance of other factors. As a result of this, and due to the large number of missing cases for this variable, income was not included in the final model.

⁵⁴⁴ Nancy Leech, Karen Barrett & George Morgan, *IBM SPSS for Intermediate Statistics: Use and Interpretation* (5th edition, Routledge 2014).

⁵⁴⁵ Agresti & Finlay (n 542).

⁵⁴⁶ David Hosmer & Stanley Lemeshow, *Applied Logistic Regression: Second Edition* (Wiley 2000).

⁵⁴⁷ UK Data Service (2014) *What is weighting?* UK Data Service <<https://www.ukdataservice.ac.uk/media/285227/weighting.pdf>> accessed 5 May 2016.

⁵⁴⁸ Goudriaan, Lynch & Nieuwbeerta (n 506).

III. Findings

(i) Bivariate analysis

(a) Socio-economic and relationship factors

Table 1. Weighted socio-economic and relationship factors

Variables	Total Sample (n = 964)	Reported to Police	p-value
Gender			.000
Male	46.8%	6.6%	
Female	53.2%	19.5%	
Age			.000
16-19	17.7%	3.0%	
20-24	17.9%	10.8%	
25-34	21.9%	16%	
35-44	21%	20.2%	
45-54	12.3%	15.6%	
55+	9.2%	15%	
NSECCA (n = 942)			.002
Managerial and professional	15.3%	14.7%	
Intermediate	9.5%	15.4%	
Supervisory and technical	8.2%	10.2%	
Routine and manual	25.3%	11.8%	
Never worked and long-term unemployed	27.6%	20.1%	
Full-time student	14.1%	4.1%	
Household income (n = 798)			.001
Above average income	47%	10.5%	
Below average income	53%	19.9%	
Scottish index of multiple deprivation (quintiles)			.069
1 (high deprivation)	28.7%	17.4%	
2	21.9%	13.9%	
3	19.8%	14.1%	
4	16.1%	9.3%	
5 (low deprivation)	13.6%	8.6%	
Whether respondent was living with partner (n = 952)			.951
Living with partner at the time	32.5%	13.6%	
Was not living with partner at the time	67.5%	13.5%	
Children (n = 954)			.000
No children	64.9%	7.8%	
Children living in household	12%	15.2%	
Children present	6.4%	21.9%	
Children involved	16.6%	32%	

Source: Scottish Crime and Justice Survey 2008-2011

Most variables relating to socio-economic status and relationship have a statistically significant effect, except: whether the respondent was living with their partner, and the Scottish index of multiple deprivation. The former was excluded from logistic regression due to the threshold of bivariate association, whereas the latter was included as it approaches significance. Those with no children are much less likely to have reported their experience to the police than those with children involved in the most recent incident (7.8% vs. 32%). Additionally, in line with existing research, a curvilinear relationship is apparent for age,⁵⁴⁹ whereby likelihood increases up to a certain point (35-44) and then decreases.

⁵⁴⁹ Akers & Kaukinen (n 499).

*(b) Consequences and nature of abuse**Table 2. Weighted reported to the police by consequences and nature of abuse*

Variables	Total Sample (n = 964)	Presents consequence	Does not present consequence	p-value
Whether respondent has ever felt a victim (n = 943)	100%	28.5%	5.7%	.000
Physical effects				
Minor bruising or black eye	24.5%	18%	12%	.020
Severe bruising	8.7%	31.9%	11.7%	.000
Scratches or minor cuts	15.5%	17.1%	12.8%	.158
Severe cuts, gashes, tears or punctures	3.6%	38.4%	12.6%	.000
Broken/cracked/fractured bones	2.3%	44.4%	12.8%	.000
Broken nose	1.2%	51.3%	13%	.000
Broken/chipped/lost teeth	1.6%	44%	13%	.000
Dislocation of joints	0.2%	35.8%	13.4%	.272
Severe concussion or loss of consciousness	1.2%	38.9%	13.2%	.013
Head injury	4.3%	38.7%	12.4%	.000
Internal injuries (such as internal bleeding or damage to internal organs)	0.7%	7.1%	13.5%	.414
Pregnancy	1.2%	44%	13.1%	.001
Contracting a sexually transmitted or other infection	0.7%	14.4%	13.5%	.942
Other physical injuries	3.6%	28.6%	12.9%	.011
Psychological effects (n = 880)				
Psychological or emotional problems	35.4%	18.9%	8.2%	.000
Fear, anxiety or panic attacks	19%	25.6%	8.8%	.000
Respondent tried to kill themselves	4.6%	38.1%	10.8%	.000
Isolation from family or friends	14%	19.1%	10.9%	.009
Feeling unable to attend work	8.6%	30.7%	10.3%	.000
Felt forced to terminate a pregnancy	1.8%	20.8%	11.9%	.273
Stopped trusting people/difficulty in other relationships	22.2%	19.2%	10%	.000
Isolation from children in their household	1.8%	26.6%	11.7%	.065
The children in their household experiencing problems or difficulties	4.3%	47.5%	10.4%	.000
Deliberately hurt themselves	5.7%	22%	11.4%	.031
Started doing things that weren't good for them to help them cope	9.6%	22%	10.9%	.003
Other negative effects other than physical injuries	6.1%	18.1%	11.6%	.142

Source: *Scottish Crime and Justice Survey 2008-2011*

Table 3. Weighted means for physical and psychological effects

Variables	Total Sample (n = 964)	Reported to the police	Not reported to the police	p-value
Number of physical effects	0.94 (0.77)	1.3 (.15)	0.6 (.04)	.000
Number of psychological effects	1.84 (.12)	2.51 (.22)	1.16 (.06)	.000

Source: *Scottish Crime and Justice Survey 2008-2011*

NOTES: Figures represent means and (standard errors)

Consequences and nature of abuse are highlighted by the literature as important, especially regarding severity. Table 2 provides evidence that differences between more and less severe consequences, as well as the majority of effects, are statistically significant. On average, victims who reported partner abuse to the police have a higher number of physical effects ($M=1.3$, $SE=.15$) than those who have not ($M=0.6$, $SE=.04$); this effect was significant. Similarly, victims reporting partner abuse have a higher number of psychological effects than those who did not, which is also statistically significant.

The effect of children in the household experiencing problems or difficulties as a result of the abuse is significant and consistent with the literature, suggesting that involvement of children is important,⁵⁵⁰ with nearly half of all respondents experiencing this consequence when reporting their victimisation. Zink and colleagues⁵⁵¹ have also noted that when mothers experiencing partner abuse noticed its effects on their children, they found greater motivation to seek change. However, many consequences have few cases; as such, only the number of physical and psychological effects were included in the regression model.

⁵⁵⁰ Akers & Kaukinen (n 499).

⁵⁵¹ Zink, Elder & Jacobson (n 502).

*(c) Confidence in the criminal justice system and the police***Table 4. Weighted confidence in the criminal justice system and the police**

Variables	Total Sample (n = 964)	Reported to Police	p-value
Confidence CJS is effective in bringing people who commit crimes to justice (n = 922)			.094
Very confident	2.7%	9.3%	
Fairly confident	44.9%	12.6%	
Not very confident	41.3%	12.8%	
Not at all confident	11.1%	21.5%	
Confidence CJS deals with cases promptly and efficiently (n = 876)			.697
Very confident	3.6%	15.8%	
Fairly confident	34.5%	12.1%	
Not very confident	41.6%	14.5%	
Not at all confident	20.4%	15.8%	
Confidence CJS provides a good standard of service for victims (n = 817)			.900
Very confident	6.5%	12.5%	
Fairly confident	44.9%	14.6%	
Not very confident	31.9%	14.5%	
Not at all confident	16.7%	16.7%	
Confidence in police's quick response (n = 928)			.090
Very confident	14.8%	17.8%	
Fairly confident	42.1%	10.3%	
Not very confident	26.4%	15.5%	
Not at all confident	16.7%	15.3%	
Confidence in police dealing with incidents as they occur (n = 933)			.395
Very confident	8.8%	17.6%	
Fairly confident	49%	12.4%	
Not very confident	27.8%	13.2%	
Not at all confident	14.4%	17.2%	
Confidence in police investigation of incidents (n = 921)			.218
Very confident	13.3%	10.2%	
Fairly confident	50%	13.7%	
Not very confident	25.4%	13.1%	
Not at all confident	11.3%	19.8%	

Source: Scottish Crime and Justice Survey 2008-2011

Table 4 presents factors relating to confidence in the CJS and the police, which were excluded from the final model because only two approach significance: confidence in CJS effectiveness in bringing people who commit crimes to justice, and confidence in local police's quick response. Although some literature identifies these as important,⁵⁵² there do not appear to be any noteworthy results, and the low count of reporting prevents analysis of all potentially relevant predictors.

⁵⁵² Bradford & Jackson (n 513).; Macdonald (n 513).

(ii) *Multivariate analysis*Table 5. *Weighted logistic regression of victim reporting of partner abuse*

Independent Variable	B	S.E.	Sig.	Exp(B)
Intercept	-4.228	.736	.000	.015
Gender				
Male	<i>Reference</i>			
Female	.738	.285	.010	2.092
Age				
16-19	<i>Reference</i>			
20-24	.232	.683	.734	1.261
25-34	1.073	.659	.104	2.925
35-44	.975	.664	.142	2.652
45-54	.619	.694	.373	1.858
55+	.293	.782	.708	1.340
NSECCA				
Managerial and professional	<i>Reference</i>			
Intermediate	.211	.465	.651	1.235
Supervisory and technical	.423	.581	.467	1.526
Routine and manual	-.196	.391	.616	.822
Never worked and long-term unemployed	.147	.369	.691	1.158
Full-time student	-.731	.678	.281	.481
Scottish index of multiple deprivation (quintiles)				
1 (high deprivation)	<i>Reference</i>			
2	-.103	.334	.759	.902
3	.145	.349	.678	1.156
4	-.341	.405	.400	.711
5 (low deprivation)	-.136	.474	.774	.873
Children				
No children	<i>Reference</i>			
Children living in household	.297	.364	.415	1.346
Children present	.747	.376	.047	2.110
Children involved	.876	.299	.004	2.400
Whether respondent has ever felt a victim				
Has not felt a victim	<i>Reference</i>			
Has felt a victim	.899	.285	.002	2.457
Number of physical effects	.262	.087	.003	1.300
Number of psychological effects	.075	.061	.217	1.078

Source: *Scottish Crime and Justice Survey 2008-2011*

Table 5 presents the results of the final regression model. Four variables are significant predictors of the probability of a victim reporting partner abuse to the police: gender, involvement of children, perceptions of victimisation, and number of physical effects. Although these measures must be used with caution, to roughly assess the fit of this model, the Cox and Snell's R^2 and Nagelkerke's R^2 were examined and returned values of .145 and .274 respectively, suggesting a good model fit.⁵⁵³

Importantly, respondents acknowledging victimisation have almost 2.5 greater odds of reporting to the police than those who did not classify themselves as a victim. Gender also plays an important role: significant positive effects were found where the victim was female; furthermore, females have two times greater odds than males of reporting. This is consistent with the literature, which suggests that men are less likely to seek help or report incidents than women.⁵⁵⁴

Also, consistent with the literature,⁵⁵⁵ was the finding that victims whose children were present or involved in the violence have over two times greater odds of reporting than those without children.

Finally, although no significant effect is apparent for the number of psychological effects, a significant effect is evident for the number of physical effects. Table 2 provides statistics on these individually. For every one-unit increase in physical effects, there is a 1.3 times increase in the odds of the victim reporting to the police. This accords with research which indicates that significant injuries increase the likelihood of reporting.⁵⁵⁶

⁵⁵³ Weisburd & Britt (n 541).

⁵⁵⁴ Patricia Tjaden & Nancy Thoennes, *Extent, nature, and consequences of intimate partner violence: Findings from the National Violence Against Women Survey*. NCJ Publication No. 181867, Washington, DC: U.S. Department of Justice, National Institute of Justice (2000) <<https://www.ncjrs.gov/pdffiles1/nij/181867.pdf>> accessed 07 May 2016.

⁵⁵⁵ Mirrlees-Black (n 499); Akers & Kaukinen (n 499); Chang, Dado, Hawker, Cluss, Buranosky, Slagel, McNeil & Scholle (n 501).

⁵⁵⁶ Bachman & Coker (n 506); Goudriaan, Lynch & Nieuwebeerta (n 506).

IV. Discussion and conclusions

This report highlights the significance of better understanding the factors which shape victim reporting to the police. In examining potential factors, it found that female victims, victims whose children were present or involved in the abuse, victims who felt victimised, and victims experiencing multiple physical effects, are most likely to report. Gaining an understanding in this area offers insight into the types of people who do and do not report violence to the police; this, in turn, has implications for various social, policy, and criminal justice issues. Only 18 per cent of those who experienced partner abuse reported it to the police; this is important since most victims are therefore not visible to the criminal justice system, which may be crucial in allowing them to leave an abusive relationship.⁵⁵⁷

It is suggested that females are more likely to report for two reasons: firstly, because of their desire for protection, and secondly, as they are less likely to perceive their partner's violence as a private matter or the incident as trivial. However, partner abuse directed at men is also a complex problem,⁵⁵⁸ and given that almost half of the sample were male victims, further examination would be useful to determine how factors differ between genders.

MacQueen and Norris⁵⁵⁹ highlighted that the effect of children's presence or involvement may result from the potential involvement of external agencies such as "health and education, or more targeted services for children and families needing greater support." However, their study examined general police awareness, with some victims not knowing how the police were informed, whereas this report usefully demonstrates the significant effect on victim reporting alone. A recent Women's Aid⁵⁶⁰ report emphasises the importance of recognising domestic abuse as harmful to

⁵⁵⁷ Akers & Kaukinen (n 499).

⁵⁵⁸ Venus Tsui, 'Male Victims of Intimate Partner Abuse: Use and Helpfulness of Services' (2014) 59 (2) Social Work 121.

⁵⁵⁹ MacQueen & Norris (n 482), 69.

⁵⁶⁰ Women's Aid, (2016) *Nineteen Child Homicides*. Bristol: Women's Aid <<https://1q7dqy2unor827bqjls0c4rn-wpengine.netdna-ssl.com/wp-content/uploads/2016/01/Child-First-Nineteen-Child-Homicides-Report.pdf>> accessed on 01 May 2016.

children, stating that domestic abuse in families with children amounts to actual harm to children even if they are not directly physically harmed. This indicates the importance of understanding the role that children play in victim reporting of partner abuse.

A central finding is that recognition of victimisation is a key factor which shapes reporting. This is important for those offering services to victims of partner abuse, including criminal justice interventions and support agencies.⁵⁶¹ Studies on victim identity have found that the term victim can negatively affect victims' propensity to seek help, particularly that from the criminal justice system because of the cultural meaning behind this term.⁵⁶² In this vein, a Spanish study demonstrated that women do not perceive all forms of abuse equally, with those subject to sexual abuse being less likely to acknowledge victimisation, perhaps due to public perceptions surrounding this social problem.⁵⁶³ The media and government also arguably contribute to the hidden nature of this problem by denying women "a voice and an opportunity to seek assistance," by failing to highlight sexual abuse as an important dimension.⁵⁶⁴ From a policy perspective, action is needed to raise awareness of partner abuse in order to encourage victims to recognise their own victimisation. Additionally, this raises issues surrounding prevention, and indicates the importance of educating the public to enable them to recognise signs of abuse.

Consistent with the literature, the finding in relation to physical effects provides further support for the argument that offence seriousness is one of the most significant predictors of reporting.⁵⁶⁵ However, contrary to expectations and MacQueen and Norris's⁵⁶⁶ recent SCJS analyses, this report found that measures of social class and deprivation, age, and variety of psychological effects did not have a significant effect on reporting when the other

⁵⁶¹ Mirrlees-Black (n 499).

⁵⁶² Jennifer Dunn, 'Accounting for Victimization: Social Constructionist Perspectives' (2008) 2 (5) 'Sociology Compass' 1601; Amy Leisenring, 'Confronting "Victim" Discourses: The Identity Work of Battered Women' (2006) 29 (3) 'Symbolic Interaction' 307.

⁵⁶³ Juanjo Medina-Ariza & Rosemary Barberet, 'Intimate Partner Violence in Spain: Findings from a National Survey' (2003) 9 (3) 'Violence Against Women' 302, 318.

⁵⁶⁴ *ibid* 318.

⁵⁶⁵ Bachman & Coker (n 506); Goudriaan, Lynch & Nieuwebeerta (n 506).

⁵⁶⁶ MacQueen & Norris (n 482).

variables in the model were controlled. Their findings also identified being a victim who had ‘never worked and long term unemployed’ to be the strongest effect in their model predicting general police awareness. However, no significant effect for this occupational category is evident for victim reporting to the police, when other factors were controlled.

Secondary data analysis is inevitably limited in terms of the variables available for analyses. Several potentially key factors which were evident in the literature, therefore, could not be included in this report. No questions were asked surrounding the use of weapons during the most recent incident and the questions relating to perceptions of law enforcement were not distinct enough to distinguish these perceptions in line with the most recent incident. Additionally, respondents were asked for their current marital status, meaning that their marital status at the time of the most recent incident is unclear.⁵⁶⁷ Ethnicity was also excluded due to disparities in the questionnaire over the three waves. Future research may benefit from the inclusion of these factors. Additionally, since this analysis discarded cases with missing data in the relevant variables and uses rough measures of the community context, future research should consider using more sophisticated analytical approaches such as multiple imputation and the use of multilevel analysis with linked census information to better capture the community context.

This research may have assumed that reporting to the police is or should be the only option for victims; however, other options are available and there are many reasons why a victim may not report to the police.⁵⁶⁸ For example, victims may be physically prevented from using a telephone,⁵⁶⁹ or by older women having links to their abuser, i.e. if they are the abuser’s carer.⁵⁷⁰ Therefore, it must be acknowledged that some may have sought informal support, for

⁵⁶⁷ *ibid* 63.

⁵⁶⁸ Akers & Kaukinen (n 499), 167.

⁵⁶⁹ Ruth Fleury, Cris Sullivan, Deborah Bybee & William Davidson, “‘Why don’t they just call the cops?’: Reasons for differential police contact among women with abusive partners’ (1998) 13 ‘Violence and Victims’ 333.

⁵⁷⁰ Therese Zink, C. Jeff Jacobson, Stephanie Pabst, Sandra Regan & Bonnie Fisher, ‘A lifetime of intimate partner violence: coping strategies of older women’ (2006) 21 ‘Journal of Interpersonal Violence’ 634.

example, through friends and family.⁵⁷¹ Furthermore, it is argued that by focusing solely on the relationship between the victims, perpetrators and the CJS, research disregards action taken within the victim's immediate social world, including those of friends, family and neighbours, in response to the violence. Consequently, it is suggested that future research should explore how communities manage knowledge of violence; for example, how and when the police are contacted.⁵⁷²

The Crime Survey for England and Wales has a self-completion module on domestic violence; however, it does not include questions surrounding police reporting. Given the importance of adequately understanding this area, the Office for National Statistics and the Home Office should reconsider the questions asked to enable analyses similar to those in this report to be undertaken.

⁵⁷¹ Catherine Kaukinen, 'The help-seeking strategies of female violent crime victims: the direct and conditional effects of race and the victim-offender relationship' (2004) 19 *Journal of Interpersonal Violence* 967.

⁵⁷² Margareta Hydén, 'What Social Networks Do in the Aftermath of Domestic Violence' (2015) 55 *British Journal of Criminology* 1040.

The regulation of the civilian and military realms of cyberspace

*Jake Taylor*⁵⁷³

This paper identifies and examines the legal predicaments with which the regulation of cyberspace is faced, whether it be civilian or military use of cyberspace. A holistic overview of cyberspace as a further realm of human activity is undertaken. A journey through this developing dominion is embarked upon, whereby consideration is given to the ways in which cyberspace can be used to commit traditional actions in a contemporary fashion, as well as how completely new types of exploits, idiosyncratic to cyberspace itself, are being realised. Specific attention is given to existing international treaties and conventions that handle the regulation of cyberspace. Upon examination of these instruments, it is clear that they fail to give sufficient guidance on the regulation of cyberspace. A shift towards clarity and predictability in the regulation of cyberspace will help in providing a safer and more secure realm.

Convolution typifies the regulation of cyberspace. Legal dilemmas are aplenty. Paranoia and hysteria accompanying the growing cyber threat run parallel to these difficulties. This is understandable, given the unclear and largely unprecedented threats that the law is expected to deal with. Cyberspace must be regulated not just to protect from potential attacks, but also for its general usage. In the civilian sphere, there are many different aspects of cyberspace that need to be considered in relation to regulation. These range from the general usage of cyberspace to cyber-based crime (cybercrime), and even common criminal offences committed through technological means. From a militaristic perspective, cyberspace offers a broad array of new challenges, particularly in relation to cyber-attacks and the use of force. This paper will focus on civilian and military applications for cyberspace and the threats therein. Real world incidents which have demonstrated the power of the cyber world will also be examined. It is important to see reactions to incidents in cyberspace on both a national level and within the international community. With an examination of the international treaties, conventions, and agreements used to regulate the cyber world, it is argued that the methods currently in place are inadequate and lack

⁵⁷³ LL.M. Candidate, The University of Manchester, School of Law.

clarity. This is particularly evident in cases of cyber warfare. Further predictability of the law in this area can help improve the safety of cyberspace for civilians and the security of cyber-based systems for the military.

Cyberspace can be understood as a “global domain within the information environment” characterised by electronic means operating through interconnected networks.⁵⁷⁴ To the layman, cyberspace is a complex world which is largely taken for granted and can never be fully understood. The multiplicity of sources from which cyberspace can now be accessed creates for a technical and vast area in which to apply legal principles. Cyberspace is no longer a concept of science fiction.⁵⁷⁵ It works hand-in-hand with the 21st-century world in which we live. From the creation of the telephone⁵⁷⁶ to the origins of the internet,⁵⁷⁷ the technological marvels of past times are now embedded in most aspects of everyday life. What is more, the rapid development of such cyber-technology means that fantastical models of the imagination are now being introduced to the public *en masse*.⁵⁷⁸ The flip-side of this progress is that its rapidity is exceeding the maturity of legal restraints or even general security controls, which are made largely on a reactive instead of proactive basis.

⁵⁷⁴ Daniel T Kuehl, ‘From cyberspace to cyberpower: Defining the problem’ in Franklin D Kramer et al, *Cyberpower and National Security* (National Defense University Press: Washington DC, 2009), p. 28.

⁵⁷⁵ ‘Cyberspace’ was coined in a science fiction novel; William Gibson, *Neuromancer* (Ace Books: New York, 1984).

⁵⁷⁶ Alexander Graham Bell is widely regarded as making the first telephone call on 10 March 1876 see, Michael E Gorman, ‘Confirmation, disconfirmation, and invention: The case of Alexander Graham Bell and the telephone’ (1995) 1 *Thinking & Reasoning* 31, p. 45.

⁵⁷⁷ The World Wide Web was created in 1989 by Tim Berners-Lee. See, Tim Berners-Lee with Mark Fischetti, *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web* (Texere: Cheshire, 2006). See also, the predecessor to the modern internet, “ARPANET” which was first imagined in 1962 see, Barry M Leiner et al, ‘A Brief History of the Internet’ (2009) 39(5) *Computer Communication Review* 22, p. 23.

⁵⁷⁸ For example; Smart Watches such as the Apple Watch (released 24 April 2015) see, <<http://www.apple.com/uk/watch/>> accessed 27 April 2016, and Virtual Reality headsets such as the Oculus Rift (released 28 March 2016) see, <<https://www.oculus.com/en-us/rift/>> accessed 27 April 2016.

I. Civilian applications, threats, and regulation in cyberspace

The world is now heavily connected. Estimates for 2015 show over seven billion mobile phone subscriptions and over three billion individuals using the internet.⁵⁷⁹ With approximations for the total world population recently hitting 7.3 billion people,⁵⁸⁰ the vast amount of technology existing globally cannot be understated. It is conceivable that mobile telephone subscriptions will soon overtake the number of people. Clearly, cyberspace has become an indelible part of the world in which we live,⁵⁸¹ and so its regulation needs to be considered. We will now briefly look at some attempts at standardisation.

In 1865, the International Telegraph Conference in Paris resulted in the first multilateral agreement of its kind to regulate the use of telegraphs.⁵⁸² The International Telegraph Union was conceived at this event. It was renamed the International Telecommunication Union (ITU) in 1932 at the International Telecommunication Conference in Madrid, reflecting the technological advancements that now fell within its remit.⁵⁸³ The ITU is the United Nations (UN) Specialized Agency focusing on information communication technologies.⁵⁸⁴ In 1992, the ITU restructured to deliver the agency we have today. This reorganisation led to three sectors of focus: Radiocommunication (ITU-R); Standardisation (ITU-T); and Development (ITU-D).⁵⁸⁵

⁵⁷⁹ International Telecommunication Union, 'Statistics', available online at, <http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx> accessed on 7 April 2016.

⁵⁸⁰ United Nations, 'World Population Prospects 2015' Population Division, Department of Economic and Social Affairs, estimates for 2015 available to download online from, <http://esa.un.org/unpd/wpp/Download/Standard/Population/> accessed 27 April 2016.

⁵⁸¹ Although generally used in the context of warfare, this "fifth domain" can also be applied generally. For an examination of the fifth domain in the context of warfare see, Jackson Maogoto, *Technology and the Law on the Use of Force: New Security Challenges in the Twenty-First Century* (Routledge: Abingdon, 2014), p. 53.

⁵⁸² Convention, Conférence Télégraphique Internationale De Paris (1865) (Hereinafter International Telegraph Convention 1865).

⁵⁸³ For an overview of the history of the ITU, see International Telecommunications Union, 'Overview of ITU's History', available online at, <http://www.itu.int/en/history/Pages/ITUHistory.aspx> accessed 7 April 2016.

⁵⁸⁴ International Telecommunication Union, 'About ITU', available online at, <http://www.itu.int/en/about/Pages/default.aspx> accessed 7 April 2016.

⁵⁸⁵ International Telecommunication Union, 'What does ITU do?', available online at, <http://www.itu.int/en/about/Pages/whatwedo.aspx> accessed 7 April 2016.

ITU-T is responsible for the handling of agreements discussed in the following paragraphs. At this point, it must be stressed that the ITU does not have any regulatory enforcement power or any goal to that end. It is the oldest intergovernmental organisation, made up of 700 members, including 193 Member States, as well as International Organisations, universities, and technology corporations.⁵⁸⁶ Much of its work is based in study groups and results of the work released through statistics, with the aim of protecting the “fundamental right to communicate.”⁵⁸⁷

The World Summit on the Information Society (WSIS) convened over two phases by the ITU – Geneva in 2003⁵⁸⁸ and Tunis in 2005⁵⁸⁹ – recognised the indispensable nature within cyberspace of the freedom of expression, flow of information, ideas, and knowledge.⁵⁹⁰ Such aims and focuses are vital to civilian enjoyment of cyberspace. As another domain of human activity, such fundamental human rights, like the freedom of expression, must be considered on an international level.⁵⁹¹ As with all international law, however, the rules pertaining to cyberspace are politically influenced. Much of the issues that can arise, do so because of traditional tensions between democratic and authoritarian standpoints. Certain agreements, such as the 2009 Agreement of the Shanghai Cooperation Organisation,⁵⁹² in which

⁵⁸⁶ International Telecommunication Union (n 584).

⁵⁸⁷ *ibid.*

⁵⁸⁸ Resulting in, World Summit on the Information Society, *Declaration of Principles. Building the Information Society: a global challenge in the new Millennium*, Document WSIS-03/GENEVA/DOC/4-E (12 December 2003) and World Summit on the Information Society, *Plan of Action*, Document WSIS-03/GENEVA/DOC/5-E (12 December 2003).

⁵⁸⁹ Resulting in, World Summit on the Information Society, *Tunis Commitment*, Document WSIS-05/TUNIS/DOC/7-E (18 November 2005) and World Summit on the Information Society, *Tunis Agenda for the Information Society*, Document WSIS-05/TUNIS/DOC/6(Rev 1)-E (18 November 2005).

⁵⁹⁰ *ibid* (*Tunis Commitment*), para. 4.

⁵⁹¹ Universal Declaration of Human Rights (adopted 10 December 1948) UNGA Res 217 A(III) (Hereinafter UDHR), Article 19. See also, International Covenant on Civil and Political Rights (adopted 16 December 1966, entered into force 23 March 1976) TS 999 171 (Hereinafter ICCPR), Article 19. See also, European Convention for the Protection of Human Rights and Fundamental Freedoms, as amended by Protocols Nos. 11 and 14 (adopted 4 November 1950, entered into force 3 September 1953) ETS5 (Hereinafter ECHR), Article 10.

⁵⁹² Agreement Among the Governments of the Member states of the Shanghai Cooperation Organisation in Cooperation in the Field of Ensuring International Information Security, 16 June 2009.

the state parties were all from an authoritarian background, demonstrate this. Definitions within the Agreement indicate a neglect of the protection of individual freedoms, such as of expression, and an emphasis on state sovereignty online.⁵⁹³ The International Treaty Regulations (ITRs) also demonstrate the opposing political tensions at work and are the focus of the next part of this essay.

Originating in 1988,⁵⁹⁴ the ITRs have been contained within a global treaty with the intention of establishing general principles with regards to international telecommunications, under the auspices of the ITU. Recommendations for amendments were most recently mooted in 2012, in which the dramatic growth of the internet ensured its position as a key topic on the agenda.⁵⁹⁵ The revisions proposed for the ITRs contained elements indicating further governance over the internet. As the ITRs were conceived in relation to telecommunications principally, the attention given to the internet seemed, to some states, as over-stretching.⁵⁹⁶ However, it would be absurd to ignore the importance of the internet in global communication. Frictions between the importance placed on protecting freedoms and protecting territorial sovereignty on the internet resulted in the unsuccessful revision of the Regulations.⁵⁹⁷ Only 89 countries, from a possible 144, signed the revised Regulations. Those who did come from a communist background and included China and Russia. Despite all of this, the Regulations do not have any authoritative legal command and hold no obligatory or mandatory power over state parties, unless they have been incorporated into national legislation. As such, the actual influence the treaty has is minimal. The ITU cannot enforce the treaty and it,

⁵⁹³ David P Fidler, 'Cyberspace and Human Rights' in Nicholas Tsagourias and Russell Buchan (eds.) *Research Handbook on International Law and Cyberspace* (Edward Elgar Publishing: Cheltenham, 2015), p. 113.

⁵⁹⁴ Final Acts of the World Administrative Telegraph and Telephone Conference, *International Telecommunication Regulations* (1988) Melbourne (WATTC-88).

⁵⁹⁵ Final Acts of the World Conference on International Telecommunications, *International Telecommunication Regulations* (2012) Dubai (WCIT-12), RESOLUTION PLEN/3, 'To foster an enabling environment for the greater growth of the Internet', p. 20.

⁵⁹⁶ n. 593 (Fidler, 2015), p. 111. See also Richard Hill, *International Telecommunication Regulations and the Internet: A Commentary and Legislative History* (Springer: Berlin, 2014)

⁵⁹⁷ *ibid.*

therefore, can only ever amount to a soft instrument,⁵⁹⁸ representing a somewhat weak grip on the regulation of civilian cyberspace by States. We will now focus on further ways in which civilian cyberspace can be regulated.

A rather idiosyncratic feature of civilian cyberspace is its layers of self-regulation. The International Corporation for Assigned Names and Numbers (ICANN) exemplifies such self-regulation. ICANN is a non-profit, private sector organisation, which maintains the online database of names and addresses, namely the Domain Name System (DNS).⁵⁹⁹ Due to the fluidity of the internet and the ease at which online pages can be formed, it is perhaps surprising that this is not regulated through a large international organisation. The plasticity of the internet is a great advantage to meeting the aforementioned freedoms, although such freedoms are where problems with self-regulation can arise.

Again, the political aspect of cyberspace for public consumption is apparent. As an American institution, ICANN adopts American principles in its work. However, as ICANN essentially regulates the 'registration plates' for the internet on a global scale, it is perplexing that this is not handled through more transnational cooperation. Through ICANN, the private sector has developed a system of self-regulation through three principal methods: the 'Code' the DNS and the 'Architecture'.⁶⁰⁰ ICANN carries out these functions through Functions Contracts with various actors, including the National Telecommunications and Information Administration (NTIA) of the United States (US) Department of Commerce, the Internet Assigned Names Authority (IANA), as well as another large US corporation, Verisign Inc.⁶⁰¹

⁵⁹⁸ Vincent G Cerf, 'The Internet at Risk' (2013) 17(2) Institute of Electrical and Electronics Engineers Computer Society 3, p. 4.

⁵⁹⁹ The Internet Corporation for Assigned Names and Numbers (ICANN), 'What Does ICANN Do?', available online at, <<https://www.icann.org/resources/pages/what-2012-02-25-en>> accessed 7 April 2016.

⁶⁰⁰ *ibid.*

⁶⁰¹ Milton L Mueller, *Networks and states: The Global Politics of Internet Governance* (MIT Press: Cambridge, 2010), p. 62-3.

Open source software further encapsulates the levels of self-regulation within civilian cyberspace. Such software, such as the Mozilla Firefox browser, can be accessed and customised by its users. The advantage of this is that the software can be adapted to the individual needs of a user, in an environment in which vulnerabilities can be recognised and resolved with speed and ease, for the benefit of the community of users operating on the same platform.⁶⁰² With these vulnerabilities comes the greater potential for abuse, however, with theoretically more widespread damage to the system as a whole in the event of such abuse, and no central support system in place.⁶⁰³

As mentioned above, the legal basis of the self-regulation that largely characterises civilian cyberspace at most only amounts to soft law.⁶⁰⁴ Nonetheless, this is not necessarily negative. Due to the tensions discussed between territorial sovereignty and individual freedoms in cyberspace, the implementation of hard law regulating civilian cyberspace has the potential to prejudice one of these positions. We will now look at some of the threats that cyberspace in the civilian domain faces.

As we have seen, there are many States, organisations, and agencies involved in working towards a cyber world which benefits all. The benefits of freedom of information and expression, however, do not come without the threats and disadvantages of such a connected world. Parallel to the advantages and expediency of cyberspace are factors that could endanger the enjoyment or even safety of those partaking in online activities. Cyber-attacks develop in the same way as other technology. As we will see, these take place against civilian and military targets across jurisdictions in a borderless online world. Although we will first deal with civilian

⁶⁰² Steven Weber, *The Success of Open Source* (Harvard University Press: Cambridge, 2004).

⁶⁰³ Greg Madey et al, 'The Open Source Software Development Phenomenon: An Analysis Based on Social Network Theory' (2002) Eighth Americas Conference on Information Systems 1806.

⁶⁰⁴ For more on Soft Organisations in international law see, Jan Klabbers, 'Institutional Ambivalence by Design: Soft Organizations in International Law' (2001) 70 *Nordic Journal of International Law* 403. For criticisms of soft law see, Jean d'Aspremont, 'Softness in International Law: A Self-Serving Quest for New Legal Materials' (2008) 19(5) *European Journal of International Law* 1075.

threats, due to the porous nature of cyberspace, the same threats can similarly affect military targets.

Attacks on computer systems can come in many forms, such as viruses, worms, trojan horses, and malicious software (malware). Infection of systems can also occur in various ways including through targeted attacks or through random targeting on a widespread basis. Cyber-attacks can cause havoc and destruction (at its worst) or inconvenience (at best) to their victims, to be considered a success for the attacker. Potential victims need to be successful in their evasion of such attacks every time. A notorious method for infiltrating systems is the 'phishing' technique. Phishing is a basic form of attack which can be extremely effective. By posing as a trustworthy source, the attacker can acquire sensitive information from its victim, if successful.⁶⁰⁵ This occurred, for example, in 2006, when the Scandinavian bank, Norden, was attacked and had sensitive financial information stolen about its customers.⁶⁰⁶

The Federal Bureau of Investigation (FBI) in the US has highlighted how serious the threats we face online can be. Cybercrime is recorded third in the list of the FBI's scale of threats, behind only terrorism and counterintelligence (espionage).⁶⁰⁷ When considering that the two threats ranked as more serious than cybercrime can also be committed through, or in, cyberspace, it makes for a harrowing prospect for future security. The mysterious world in which cybercriminals operate has led to the inception of a somewhat 21st-century folk devil in the form of online criminals and hackers.⁶⁰⁸

⁶⁰⁵ Rachna Dhamija, J D Tygar and Marti Hearst, 'Why Phishing Works' (2006) *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery 581. For a numeric demonstration in relation to spam and phishing see, Cyberroam and Commtouch, 'Internet Threats Trend Report April 2012' (2012) Q1, available online at, <http://www.cyberoam.com/downloads/ThreatReports/Q12012InternetThreats.pdf> accessed 7 April 2016.

⁶⁰⁶ Roland Heckerö, *The Dark Side of the Internet: On Cyber Threats and Information Warfare*, Martin Peterson (trs.) (Peter Lang GmbH Internationaler Verlag der Wissenschaften: Frankfurt, 2013), p. 101.

⁶⁰⁷ Federal Bureau of Investigation, 'What we investigate', available online at, https://www.fbi.gov/about-us/investigate/what_we_investigate accessed 27 April 2016.

⁶⁰⁸ For more on 'folk devils' see, Stanley Cohen, *Folk Devils and Moral Panics* (MacGibbon and Kee Ltd: London, 1972).

In the face of a rising cybercrime problem, the Cybercrime Convention of 2001 was adopted.⁶⁰⁹ It was a landmark in that it was the first multilateral treaty directly concerned with crimes committed online. The Convention dealt with various issues regarding online criminality, with a focus on computer-related fraud,⁶¹⁰ offences related to child pornography,⁶¹¹ and copyright violations.⁶¹² Much of the foci were therefore on offences not idiosyncratic to cyberspace. However, it still marked an important step in the international acknowledgement of the criminal activity that can take place in cyberspace. International agreement on the principles contained within the Convention had the goal of percolating into national legislation and thus becoming hard law.⁶¹³ Conventions such as these are necessary for the development of offences to be applied to acts committed in cyberspace.

Another way in which the cyber world can have ramifications in the physical world is through the ‘deep web’. and more specifically, the ‘dark net’. The deep web is the label given to the unindexed content that sits beneath the visible content on the internet (the ‘surface web’).⁶¹⁴ The dark net is the hidden area of the deep web which requires software, such as the military-grade ‘Tor’, to gain access.⁶¹⁵ As its appellation suggests, the dark net is hidden. Use of it cannot be traced by law enforcement, intelligence, or any other government agencies (or anyone else for that matter) in the way that actions on the conventional internet can be. This makes it a popular environment for criminals to operate, as well as journalists and even undercover police officers.⁶¹⁶ As with much of the

⁶⁰⁹ Council of Europe Convention for Cybercrime (adopted 23 November 2001, entered into force 1 July 2004) ETS 185 (Hereinafter Convention for Cybercrime).

⁶¹⁰ *ibid*, Article 8.

⁶¹¹ *ibid*, Article 9.

⁶¹² *ibid*, Article 10.

⁶¹³ For example, in the United Kingdom, it came into force through, Convention for Cybercrime (1 September 2011) Treaty Series No. 18 (2012) Cmnd 3809.

⁶¹⁴ Hsinchun Chen, *Dark Web: Exploring and Data Mining the Dark Side of the Web* (Springer: New York 2012).

⁶¹⁵ Roger Dingledine, Nick Mathewson and Paul Syverson, ‘Tor: The Second-Generation Onion Router’ (2004) Washington DC: US Naval Research Lab. See also, ‘Tor’, available for download at <<https://www.torproject.org>> accessed 27 April 2016.

⁶¹⁶ David Omand, ‘The Dark Net: Policing the Internet’s Underworld’ (2015) 32(4) World Policy Journal 75. See also generally, Alex Winter (Dir.), *Deep Web* [Documentary] (Epix, Paramount Pictures: New York, 2015).

technology discussed already, the effect of the dark net is two-fold and has both advantages and disadvantages to its actors.

The infamous ‘Silk Road’ site was at one point the most popular marketplace on the dark net, in which members could buy and sell products and services that could not be traded on the conventional internet. To clarify, Silk Road made headlines for the ability of its users to purchase drugs, and even, for example, order the services of a hitman from their computer in a virtually untraceable transaction.⁶¹⁷ It is reported that it made \$1.2 billion in sales,⁶¹⁸ all underpinned by the virtual currency ‘Bitcoin’, further adding to the covert nature of the medium. After extensive investigations and legal processes, the FBI arrested the supposed mastermind of Silk Road. 29-year-old Ross Ulbricht was arrested whilst logged into the administrator’s account in a public library in San Francisco.⁶¹⁹ He was eventually convicted and received five sentences including two for life imprisonment.⁶²⁰ The full force of the law, usually used in the physical world, was felt by Ulbricht because of actions taken in cyberspace.

What is particularly intriguing about the Silk Road saga is the way in which the FBI managed to shut down the website. They traced the servers to Iceland and Germany,⁶²¹ but the techniques used to find them have not been revealed. This has led to accusations of the FBI hacking into Silk Road,⁶²² that is, fighting criminal enterprise through a violation of individual rights (in this case, Ross

⁶¹⁷ Martin James, ‘Lost on the *Silk Road*: Online drug distribution and the “cryptomarket”’ (2014) 14(3) *Criminology & Criminal Justice* 351. See also generally, Alex Winter (Dir.), *Deep Web* [Documentary] (Epix, Paramount Pictures: New York, 2015).

⁶¹⁸ Adrienne Jeffries, ‘FBI seizes underground drug market Silk Road, owner indicted in New York’ (2013) *The Verge*, available online at, <<http://www.theverge.com/2013/10/2/4794780/fbi-seizes-underground-drug-market-silk-road-owner-indicted-in-new>> accessed 20 April 2016.

⁶¹⁹ Thielman Sam, ‘Silk Road operator Ross Ulbricht sentenced to life in prison, *The Guardian*, (London 29 May 2015), available online at, <<https://www.theguardian.com/technology/2015/may/29/silk-road-ross-ulbricht-sentenced>> accessed 20 April 2016.

⁶²⁰ *United States of America v. Ross William Ulbricht*, 14 Cr. 68 (KBF), US District Court, New York, 26 May 2015.

⁶²¹ Andy Greenberg, ‘FBI’s Story of Finding Silk Road’s Server Sounds a Lot Like Hacking’ (2014) *Wired*, available online at, <<https://www.wired.com/2014/09/fbi-silk-road-hacking-question/>> accessed 7 April 2016. See also, Alex Winter (Dir.), *Deep Web* [Documentary] (Epix, Paramount Pictures: New York, 2015).

⁶²² *ibid.*

Ulbricht's). Legal dangers for the authorities themselves then become somewhat apparent. In this case, the issue at hand was over rights contained within the US Constitution relating to improperly obtaining evidence.⁶²³ Similar principles can also be found in international legal instruments, such as comparable articles in the Universal Declaration of Human Rights (UDHR).⁶²⁴ If law enforcement agencies turn to hacking online communities and individuals within them, it sets a dangerous precedent and runs the risk of encroaching on the rights and freedoms of individuals. The case also demonstrates a lack of international cooperation in the handling of cybercrime perpetrators.

Actions in cyberspace can also result in legal consequences in the physical world on a smaller scale. This can be seen through abusive and deliberately inflammatory comments made online, often anonymously, which has been termed 'trolling'.⁶²⁵ For example, in the United Kingdom (UK), the Malicious Communications Act can and has been used, to prosecute individuals for such online abuse.⁶²⁶

We have seen how actions in cyberspace can have effects in the physical realm regarding civilians. Such actions can also have a simultaneous effect in cyberspace. For example, online trolling on social media sites such as Facebook⁶²⁷ and Twitter⁶²⁸ can lead to online sanctions. Blocking or removing particular accounts from such sites is only a short-term solution and has the potential to become somewhat circular, again due to the plasticity of the internet. Certain actions in cyberspace can result in more destructive effects, for example economically, as well as personal ramifications, for victims on an individual level. Recent history provides a wealth of examples in which this has occurred. Across 2014 and 2015 alone,

⁶²³ Constitution of the United States, Amendment IV.

⁶²⁴ n. 591, UDHR, Article 12. See also n. 591, ICCPR, Articles 17. See also, n. 591, ECHR, Article 8.

⁶²⁵ Oxford English Dictionary, [Draft Addition in March 2006].

⁶²⁶ Malicious Communications Act 1988, Section 1, as amended by the Criminal Justice and Courts Act 2015.

⁶²⁷ <www.facebook.com>.

⁶²⁸ <www.twitter.com>.

some high-profile attacks include those on Sony Pictures,⁶²⁹ the targeted 'Dark Hotel' malware,⁶³⁰ and the notoriously widespread hacking of personal details of the users of adultery website Ashley Madison.⁶³¹ Attacks have also occurred on high street banks,⁶³² software downloaded from the internet,⁶³³ and on credit card payment systems.⁶³⁴ As stated above, the security in place for such attacks is developed on a reactive basis. Additionally, striving for justice for cyber-attacks is often futile due to the inherent anonymity of cyberspace.

Cybersecurity on a state level presents another platform in which political attitudes can have influence. This can be observed through, for example, an authoritarian state such as China. Cybersecurity has different connotations for States such as China, as, through online censorship of its residents, control over challenges to the authority of the communist regime can be achieved.⁶³⁵ We will now look at how cyberspace is used in a military setting, with attention given to real-world incidents that have shaped the discussion around regulation of cyberspace for military purposes.

⁶²⁹ Lisa Ricchwine, 'Cyber attack could cost Sony studio as much as \$100 million' (2014) Reuters, available online at, <<http://uk.reuters.com/article/us-sony-cybersecurity-costs-idUSKBN0JN2L020141209>> accessed 1 May 2016.

⁶³⁰ Kim Zetter, 'DarkHotel: A Sophisticated New Hacking Attack Targets High-Profile Hotel Guests' (2014) Wired, available online at, <<https://www.wired.com/2014/11/darkhotel-malware/>> accessed 1 May 2016.

⁶³¹ Robert Hackett, 'What to know about the Ashley Madison hack' (2015) Fortune, available online at, <<http://fortune.com/2015/08/26/ashley-madison-hack/>> accessed 15 April 2016.

⁶³² For example, Richard Adhikari, 'RBS WorldPay Data Breach hits 1.5 million' internetnews.com, 24 December 2008, available online at, <<http://www.internetnews.com/security/article.php/3793386/RBS+WorldPay+Data+Breach+Hits+15+Million.htm>> accessed 27 April 2016.

⁶³³ Such as the infamous malware affecting Adobe software. See, Alexandre Blonce et al, 'Portable Document Format (PDF) Security Analysis and Malware Threats' (2008) Virology and Cryptology Academy.

⁶³⁴ Thomas Claburn, 'Heartland Payment System Hit by Data Security Breach' (2009) Information Week, available online at, <<http://www.darkreading.com/attacks-and-breaches/heartland-payment-systems-hit-by-data-security-breach/d/d-id/1075770?>> accessed 27 April 2016.

⁶³⁵ Nigel Inkster, 'China in cyberspace' (2010) 52(4) Survival 55.

II. Military applications, threats, and regulation in cyberspace

The military sphere relies on technology and cyberspace to a similar extent as the civilian world. For example, military operations on the ground are supported with technologies from radio contact to satellite imagery support. In addition to traditional methods of warfare in which technology plays its part, the “fifth domain” of warfare is indeed cyberspace.⁶³⁶ As already indicated above, this follows the traditional methods of land, sea, and air, as well as the possibility of warfare in the fourth domain of outer space.⁶³⁷ Throughout the last decade, there have been notable attacks orchestrated by, and directed at, States and their online systems, such as ‘Code Red’,⁶³⁸ ‘Turla’,⁶³⁹ ‘Regin’,⁶⁴⁰ and the work of the ‘Equation Group’.⁶⁴¹ Again, attribution of such attacks is an issue. The law that governs the fifth domain of cyberspace in a military context will be the focus of this part of the paper.

How cyber-attacks directed towards States are perceived by the international community is decisive regarding the methods used to deal with them. Such attacks can be distinguished as Information Operations (IO) which generally occur outside the existence of a conflict, and Information Warfare (IW) which is resorted to during a conflict.⁶⁴² In short, a major issue facing international law is

⁶³⁶ n. 581 (Maogoto, 2014), p. 53. See also, Toby L Friesen, ‘Resolving Tomorrow’s Conflicts Today: How New Developments Within the U.N. Security Council can be Used to Combat Cyber Warfare’ (2009) 58 Naval Law Review 89.

⁶³⁷ *ibid* (Maogoto, 2014), p. 31. See also, United States Space Command, *Long Range Plan: Implementing USSPACECOM Vision for 2020*, (Peterson Air Force Base: Colorado, 1998).

⁶³⁸ Nicole C Wong, “‘Code Red’ Creeping Worldwide”, *The Washington Post* (Washington D.C. 2 August 2001), available online at, <<http://www.washingtonpost.com/wp-srv/post/intern/stories01-13.htm>> accessed 20 April 2016.

⁶³⁹ Elizabeth Hanford, ‘The Cold War of Cyber Espionage’ (2014) 20(1) Public Interest Law Reporter 22, p. 22.

⁶⁴⁰ For more see, Wojciech Mazurczyk and Luca Caviglione, ‘Information Hiding as a Challenge for Malware Detection’ (2015) 13(2) IEEE Security and Privacy Magazine 89.

⁶⁴¹ Kaspersky Lab, ‘Equation Group: The Crown Creator of Cyber-Espionage’, available online at, <<http://www.kaspersky.com/about/news/virus/2015/equation-group-the-crown-creator-of-cyber-espionage>> accessed 7 April 2016.

⁶⁴² Michael Schmitt, ‘Computer Network Attack and the Use of Force in International Law: Thoughts on a Normative Framework’ (1999) 37 Columbia Journal of Transnational Law 885, p. 891. See also, James C Duncan, ‘A Primer on the Employment of Non-Lethal Weapons’ (1998) 45 Naval Law Review 1, p. 7. These are not new conceptions, but have new means and methods for execution see, n. 606 (Heickerö, 2013), p. 31.

whether cyber-attacks are synonymous to a use of force in the kinetic world, as prohibited by the UN Charter.⁶⁴³ Generally, and historically, this has been considered to be the use of physical force, such as invasions or the use of physical weaponry against other States, outside the confines of an armed conflict.⁶⁴⁴ Indeed, with the use of force in the physical world, States have an “inherent” right to self-defence.⁶⁴⁵ If cyber-attacks are to be viewed in the same vein as kinetic attacks, the question of whether self-defence can be invoked becomes relevant.

Estonia, with its 1.3 million inhabitants,⁶⁴⁶ is one of the most technologically connected nations in the world today. Much of its national infrastructure is based upon and within technological means. For example, 97% of all banking transactions in Estonia are made online.⁶⁴⁷ Whilst this delivers convenience, it inevitably leaves the country vulnerable to attacks on the structure upon which it is based. Such attacks can be committed from anywhere else with internet access; this vulnerability was indeed exploited in 2007, in the first cyber-attack of its kind against a sovereign state.⁶⁴⁸ Following the movement of the ‘Monument to the Liberators of Tallinn’ statue from the city centre in Tallinn to a military cemetery, riots and cyber-attacks against Estonia ensued. The monument was originally built in 1944 during the Soviet era, and its relocation angered many Russians. Through two main waves of attack, Estonia

⁶⁴³ Charter of the United Nations (adopted 26 June 1945, entered into force 24 October 1945) 1 UNTS XVI (Hereinafter UN Charter), Article 2(4). ‘Force’ in Article 2(4) has been seen as broad enough to include cyber-attacks see, Richard W Aldrich, ‘How do you know you are at war in the information age?’ (2000) 22 *Houston Journal of International Law* 223, p. 254.

⁶⁴⁴ The type of force is not specified in Article 2(4) UN Charter see, Ian Brownlie, *International Law and the Use of Force by states* (Clarendon Press: Oxford, 1963) p. 265-70. See also generally, Todd A Morth, ‘Considering Our Position: Viewing Information Warfare as a Use of Force Prohibited by Article 2 (4) of the UN Charter’ (1998) 30 *Case Western Reserve Journal of International Law* 567.

⁶⁴⁵ n. 643 (UN Charter), Article 51.

⁶⁴⁶ n. 580.

⁶⁴⁷ n. 606 (Heickerö, 2013), p. 134.

⁶⁴⁸ Ian Traynor, ‘Web Attackers Used a Million Computers, Says Estonia’, *The Guardian*, (London 18 May 2007), available online at, <<https://www.theguardian.com/technology/2007/may/18/news.russia>> accessed 15 April 2016. See also, Wyatt Kash, ‘Lessons from the cyberattacks on Estonia’ (2008) *Government Computer News*, (Interview with Lauri Allman, Estonia’s Permanent Undersecretary of Defence), available online at, <<https://gcn.com/articles/2008/06/13/lauri-almann--lessons-from-the-cyberattacks-on-estonia.aspx>> accessed 7 April 2016.

was victim to Distributed Denial of Service (DDoS) assaults on a scale that had not previously been witnessed.⁶⁴⁹ The first wave was less sophisticated than the second, which used large ‘botnets’ of compromised computers to launch many DDoS attacks against the state.⁶⁵⁰ This resulted in much of the technological functioning of the state coming to a standstill. Of course, this had a negative impact on Estonia, particularly from an economic standpoint.⁶⁵¹

Although there has never been any conclusive evidence to this effect, it is strongly believed that the cyber-attacks emanated from Russia, and were committed by Russians.⁶⁵² As Estonia is a NATO member state, collective self-defence could be invoked under the North Atlantic Treaty in the event of an armed attack.⁶⁵³ In the wake of the cyber-attacks on Estonia, the incident was dealt with entirely under criminal law and no invocation of collective self-defence was made.⁶⁵⁴ Such action, or inaction, may indicate that the international perspective on cyber-attacks is that they do not reach the threshold of an armed attack.⁶⁵⁵ On the other hand, as this area of law is clearly in its infancy, the international response may be more subdued and restrained. The unprecedented nature of the

⁶⁴⁹ *ibid* (Traynor, 2007).

⁶⁵⁰ n. 606 (Heickerö, 2013), p. 132.

⁶⁵¹ It is generally considered that economic coercion is not prohibited by Article 2(4) UN Charter. See, Albrecht Randelzhofer, ‘Article 2 (4)’ in Bruno Simma et al (eds.) *The Charter of the United Nations: A Commentary*, (2nd edn, Oxford University Press: Oxford, 2002), p. 118. For an example of high-level targeted economic cyber-attacks see the Shamoon virus, Symantec Security Response, ‘The Shamoon Attacks’ (2012) Symantec Official, available online at, <<http://www.symantec.com/connect/blogs/shamoon-attacks>> accessed 7 April 2016.

⁶⁵² Thais Portilho-Shrimpton, ‘Battle for South Ossetia Fought in Cyberspace’ *The Independent*, (London 17 August 2008) available online at, <<http://www.independent.co.uk/news/world/europe/battle-for-south-ossetia-fought-in-cyberspace-899772.html>> accessed 27 April 2016.

⁶⁵³ North Atlantic Treaty (adopted 4 April 1949, entered into force 24 August 1949) 34 UNTS 243 (Hereinafter North Atlantic Treaty), Article 5. See also n. 643 (UN Charter), Article 51.

⁶⁵⁴ Sverre Myrli, ‘173 DSCFC 09 E BIS - NATO and Cyber Defence’ (2009) NATO Parliamentary Assembly, paras 58-61, available online at, <<http://www.nato-pa.int/default.asp?SHORTCUT=1782>> accessed 20 April 2016. See also, Heather Harrison Dinniss, *Cyber Warfare and the Laws of War* (Cambridge University Press: Cambridge, 2012), p. 39.

⁶⁵⁵ This is the position taken by Professor Tsagourias in, Nicholas Tsagourias, ‘Cyber attacks, self-defence and the problem of attribution’ (2012) 17 (2) *Journal of Conflict and Security Law* 229, p. 232.

attack may have highlighted an international community which did not know how to respond.

Another incident of a similar scale took place during the 2008 conflict in Georgia. In another prime example of unprecedented events taking place, the world witnessed the first combination of cyber-attacks with a military offensive.⁶⁵⁶ Before any military action had begun, DDoS attacks on a small scale were launched against the state.⁶⁵⁷ Georgia suffered website defacements and more seriously, larger DDoS attacks, as well as 'SQL injections', which supplement DDoS attacks through installing junk code which is intended to mystify the back-end database of the websites being targeted.⁶⁵⁸ As with the cyber-attacks on Estonia, although Russia was, and is, suspected of having carried out the attacks, there is no concrete evidence to link the Russian authorities to the events in Georgia.⁶⁵⁹ This highlights the recurring issue of attribution, which will be developed further, later in this paper.

The introduction, also in 2008, of the largest computer worm known to date, went on to infect more computers than any other attack.⁶⁶⁰ 'Conficker', as it is known, was specially designed to attack computers running a Windows operating system.⁶⁶¹ The Conficker Worm again illustrates the issue of attribution. It is still unclear who is, or was, behind the creation of the Worm.⁶⁶² This impedes the development of legal doctrine dealing with the regulation of cyberspace. If the actions of individuals, organisations, and States cannot be accurately attributed to them, an uncertain future awaits, with levels of paranoia within the international community that have not been witnessed since the Cold War era.

⁶⁵⁶ Council of the European Union, Independent International Fact-Finding Mission on the Conflict in Georgia, 2009 Report, (Vol II, September 2009).

⁶⁵⁷ Jose Nazario and Andre Dimino, 'An In-Depth Look at the Georgia-Russia Cyber Conflict of 2008' Arbour Networks, Shadowserver Foundation, BTF8, available online at, <https://www.shadowserver.org/wiki/uploads/Shadowserver/BTF8_RU_GE_DDOS.pdf> accessed 1 April 2016.

⁶⁵⁸ n. 606 (Heickerö, 2013), p. 138.

⁶⁵⁹ n. 652 (Portilho-Shrimpton, 2008).

⁶⁶⁰ For statistics on Conficker, see, Shadowserver, 'Conficker', available online at, <<http://www.shadowserver.org/wiki/pmwiki.php/Stats/Conficker>> accessed 1 April 2016.

⁶⁶¹ n. 606 (Heickerö, 2013).

⁶⁶² *ibid.*, p. 100.

Perhaps inevitably, the advent of the Stuxnet worm demonstrated the potential for a major cyber-attack. In the first experience of its kind, the Iranian Nuclear Program was infiltrated by a computer-born weapon in 2009, targeted at disrupting the nuclear centrifuges in the Natanz uranium enrichment facility.⁶⁶³ It is the first worm to be credited with targeting real-world infrastructure.⁶⁶⁴ The potential of such cyber warfare weapons was immediately realised. This worm had penetrated a security system of the highest level and could directly manipulate the nuclear materials the system was meant to protect. Disaster could plausibly have ensued. What further complicates matters is that the attack appeared to be almost certainly the actions of States, due to the level of sophistication and the strength and capability of the worm.⁶⁶⁵ Israel and the US have become the main suspects for the Stuxnet attack. It almost does not matter whether they were responsible for its creation or not. The fact that this had occurred and international law had no recourse to deal with the attack is the main concern. There is growing trepidation across the international community towards the potential for terrorist groups to gain control of nuclear materials and make a weapon with such capabilities.⁶⁶⁶ A targeted cyber-attack in the vein of Stuxnet is indeed a plausible technique by which such groups could gain control of dangerous materials.

Computer viruses have not always been of such high levels of sophistication. The first viruses to emerge, in the 1980s, mainly *destroyed* information on the computers being targeted rather than aiming to *steal* that information.⁶⁶⁷ Such malware has developed into the highly intricate labyrinths of code in action today. This has all happened over a remarkably short space of time. To put it into

⁶⁶³ Marco Roscini, 'Cyber operations as a use of force' in Nicholas Tsagourias and Russell Buchan (eds.) *Research Handbook on International Law and Cyberspace* (Edward Elgar Publishing: Cheltenham, 2015), p. 243.

⁶⁶⁴ n. 581 (Maogoto, 2014), p. 53.

⁶⁶⁵ M Moore, 'War with Iran? Stuxnet may be first Cybersalvo' (2010) *Pacific Standard*, available online at, <<https://psmag.com/war-with-iran-stuxnet-may-be-first-cybersalvo-94f40d7c6663#.qdc6ye8wz>> accessed 1 April 2016.

⁶⁶⁶ Terrorist websites have been noted as detailing how to launch DDoS attacks, for example see, Larry Greenemeir, 'Electronic Jihad App Offers Cyber Terrorism For the Masses' (2007) *Information Week*, available online at, <<http://www.informationweek.com/electronic-jihad-app-offers-cyberterrorism-for-the-masses/d/d-id/1056683?>> accessed 27 April 2016.

⁶⁶⁷ n. 606 (Heickerö, 2013), p. 95.

perspective, contemporary weapons on the battlefield have developed from the Chinese creation of gunpowder in the ninth century, into the various forms of contemporary artillery.⁶⁶⁸ In stark contrast, the internet has only been available for public consumption since the latter half of the 20th century.⁶⁶⁹ This relates to the ‘Revolution in Military Affairs’ (RMA) the world is currently witnessing, in which a transformation in the way opposing sides are conducting wartime activities is taking place.⁶⁷⁰ Indeed, major States including the US, the UK, and China, now have dedicated cyber commands embedded within their military.⁶⁷¹

It is certainly a very telling time for international law. Law, particularly international law, which is largely based on custom,⁶⁷² faces perhaps its biggest challenge yet. The technology behind potential cyber warfare is advancing at an astounding rate. Even the *technological* defences to cyber-attacks are left in the dust of offensive capabilities. For laws to be developed in such an environment is a monumental challenge. New treaties, conventions, and agreements would realistically take many years of ripening to tackle the issues. However, applying existing international legal principles to these new developments carries its own demands. As the world scratches its head over the response to such events, the

⁶⁶⁸ Jared Diamond, *Guns, Germs and Steel: A Short History of everybody for the last 13,000 years* (Vintage Books: London, 1998), p. 253. Also, see generally, John Ellis, *The Social History of the Machine Gun* (2nd edn, John Hopkins University Press: Baltimore, 1986).

⁶⁶⁹ n. 577 (Berners-Lee with Fischetti, 2006), p. 124.

⁶⁷⁰ Michael Schmitt, ‘Wired warfare: computer network attack and the *jus in bello*’ (2002) in Schmitt and O’Donnell, *Computer Network Attack and International Law*, 76 *Naval War College International Law Studies* 365, p. 365. See also, Matthew Mowthorpe, ‘The Revolution in Military Affairs (RMA): The United States, Russian and Chinese Views’ (2005) 30(2) *The Journal of Social, Political and Economic Studies* 137. See also, n. 581 (Maogoto, 2014), p. 23. For a discussion of attacks to opposing military information infrastructure see, Brian T O’Donnell and James C Kraska, ‘Humanitarian Law: Developing International Rules for the Digital Battlefield’ (2003) 8(1) *Journal of Conflict and Security* 133, p. 138.

⁶⁷¹ n. 654 (Dinniss, 2012), p. 53. For samples of approaches around the world see, Jason Andress and Steve Winterfeld, *The Basics of Cyber Warfare: Understanding the Fundamentals of Cyber Warfare in Theory and Practice* (Syngress: Waltham, 2013), p. 40-50.

⁶⁷² Statute of the International Court of Justice (adopted 26 June 1945, entered into force 24 October 1945) 33 TS 993 (Hereinafter ICJ Statute), Article 38(1)(b). The proposition of ‘instant’ custom has been made see, Bin Cheng, ‘United Nations Resolutions on Outer Space: “Instant” International Customary Law?’ (1965) 5 *Indian Journal of International Law* 23, but this is generally rejected.

shady area of little to no regulation is being evidenced by viruses and worms such as Stuxnet. The deniability that comes hand in hand with cyber-attacks is part of the reason for this. But the lack of proper regulation using established international legal principles means that when serious cyber-attacks from States on other States occur, there may be no obvious redress that can be sought by the victim state. We will now examine the options available in such circumstances.

Much of the legal framework used to regulate cyberspace comes from existing treaties and custom surrounding, for example, armed attacks. When such treaties dealing with armed attacks were first drafted, the only consideration at the time was of military force.⁶⁷³ The biggest obstacle for states and international organisations, progressing into the technological age has been, and will be, the use of these existing laws, to suit an area of law which is evolving at a much slower rate than the technology for which it is meant to regulate.

Use of force may no longer automatically mean the use of traditional artillery on the battlefield. Indeed, it is unlikely that anyone would argue that the use of a nuclear weapon in an attack against another state would not amount to a use of force. This is just another development of technology.⁶⁷⁴ If cyber-attacks can cause physical damage in the ‘real’ world then there is no legitimate argument as to why it should not be considered a use of force.⁶⁷⁵

⁶⁷³ Elizabeth Wilmshurst, ‘Principles of International Law on the Use of Force by states in Self-Defence’ (2005) Chatham House, ILP WP 05/01, The Royal Institute of International Affairs, p. 6. See also, regarding aggression, United Nations General Assembly ‘Definition of Aggression’ (14 December 1974) UN Doc A/RES/3314(XXIX) (Hereinafter UNGA Res 3314), Article 1.

⁶⁷⁴ n. 581 (Maogoto, 2014), p. 14. See also, Scott Shackelford, ‘From Nuclear War to Net War: Analogizing Cyber Attacks in International Law’ (2009) 27(1) Berkeley Journal of International Law 192.

⁶⁷⁵ See the ‘scale and effect’ test employed by the ICJ in *Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. United States of America)*, Merits, 27 June 1986, ICJ Rep, p. 14 (Hereinafter *Nicaragua Case*), para. 195. See also, Michael Schmitt (ed.) *Tallinn Manual on the International Law Applicable to Cyber Warfare* (Cambridge University Press: Cambridge, 2013), Rule 11, p. 45-52. See also, Yoram Dinstein, ‘Computer Network Attacks and Self-Defense’ (2002) in Michael Schmitt and Brian T O’Donnell (eds.) *Computer Network Attack and International Law*, 76 International Law Studies 99, p. 103. See also Harold Koh, ‘International Law in Cyberspace’ (2012) Faculty Scholarship Series, Paper 4854, 54 Harvard International Law Journal, p. 4.

Demonstrating the perspective towards cyber-attacks that they adopt, Russia has previously adopted a policy in which they could essentially respond with weapons of mass destruction to a cyber-attack against the state.⁶⁷⁶ Strong and intimidating rhetoric indeed, and whilst sounding slightly preposterous that this would genuinely come into consideration, it signifies the attitude towards such attacks, the paranoia surrounding them, and the severity with which they are being treated.

Indeed, the case can be made that the issue at hand does not concern the weapons used to administer force on a target, but the actual force that is administered.⁶⁷⁷ From this consequentialist perspective, if it can be ascertained that the effect of a cyber-attack amounts to the use of force, then the international frameworks already in place can be applied to such situations. What must be clear is the threshold of a use of force. This is where lies many of the problems and disagreements in this field.⁶⁷⁸

The threshold for a use of force (armed attack) with which a cyber-attack must reach can be seen through different perspectives. One viewpoint is offered by Professor Reisman, who amalgamates several categories that represent force, regardless of legality, as having occurred if coercive actions challenge or enrich world order.⁶⁷⁹ This can be understood in the light of a target-based approach and applies irrespective of damage done. If an attack is aimed at national infrastructure, it will automatically be considered

⁶⁷⁶ Martin Libicki, *Cyberdeterrence and cyberwar* (The RAND Corporation: California, 2009), p. 69. The US have also said they will react militarily to cyber-attacks see, Dan Verton, 'The Prospect of Iraq Conflict Raises New Cyber Attack Fears' (2002) Computerworld, available online at, <<http://www.computerworld.com/article/2579191/cybercrime-hacking/prospect-of-iraq-conflict-raises-new-cyberattack-fears.html>> accessed 20 April 2016.

⁶⁷⁷ *Legality of the Threat or Use of Nuclear Weapons*, Advisory Opinion, 8 July 1996, ICJ Rep, p.226 (Hereinafter Nuclear Weapons Advisory Opinion), para. 39. Armed attack can also take an indirect form at times see, n. 102 (*Nicaragua* Case), paras. 93-9. For a discussion on the instrument, effects and target approaches see, n. 663 (Roscini, 2015), p. 237. See also, Matthew J Sklerov, 'Solving the Dilemma of State Responses to Cyberattacks: A Justification for the Use of Active Defences Against states Who Neglect Their Duty to Prevent' (2009), 201 Military Law Review 1.

⁶⁷⁸ n. 663 (Roscini, 2015), p. 235.

⁶⁷⁹ Michael Reisman, 'Criteria for the Use of Force in International Law' (1985) 10 Yale Journal of International Law 279, p. 281-2. See also, Christopher C Joyner and Catherine Lotrionte, 'Information Warfare as International Coercion: Elements of a Legal Framework' (2001) 12(5) European Journal of International Law 825, p. 846.

a use of force. The target-based approach is over-inclusive, however.⁶⁸⁰ If a small-level attack which did little or no damage to critical national infrastructure was worthy of being perceived as a use of force, a dangerous precedent might be set. On the other hand, coercion and violence on a large and international scale could amount to the use of force, through the lens of the New Haven school.⁶⁸¹ This can be more closely assimilated with a ‘consequence’ or ‘effects-based’ approach. Through this approach, the cyber-attack will only reach the threshold of an armed attack if it has direct damaging effects on people and objects. Cyber-based attacks can target many facets of infrastructure, which all have the potential to have such damaging effects on people and objects.⁶⁸²

If states were to respond to cyber-attacks, under their inherent right to self-defence, as embodied in the UN Charter,⁶⁸³ there are attributional issues that must be overcome. Due to the nature of cyber-attacks and cyberspace in general, this may take considerably longer and involve much more complex processes than with traditional modes of conflict. Attribution will require technical and political calculation on top of the necessary legal attribution.⁶⁸⁴ Cyberspace offers attackers anonymity, the platform to launch multiple, simultaneous, and successive attacks, as well as the opportunity to deliver attacks with a haste that is alien to comparable kinetic attacks.⁶⁸⁵

Increasingly, non-state actors are becoming involved in attacks launched against both civilian and state targets. Vigilante organisations such as the infamous, and satirically named, hacker group ‘Anonymous’ have progressively claimed responsibility for high-profile actions online.⁶⁸⁶ Other groups which actually embody

⁶⁸⁰ n. 663 (Roscini, 2015), p. 236.

⁶⁸¹ Myers S McDougal and Florentino P Feliciano, *Law and Minimum World Public Order: The Legal Regulation of International Coercion* (Yale University Press: New Haven, 1961).

⁶⁸² It has even been suggested that if a cyber-attack was directed at a State’s national infrastructure such as the financial system as a whole, it could amount to a use of force see, n. 655 (Tsagourias, 2012), p. 232.

⁶⁸³ n. 643 (UN Charter), Article 51.

⁶⁸⁴ n. 655 (Tsagourias, 2012).

⁶⁸⁵ *ibid.*, p. 234.

⁶⁸⁶ For example, their efforts against the so called Islamic State. See, Alex Hern, ‘Anonymous “at war” with Isis, hacktivist group confirms’, *The Guardian* (London 17 November 2015),

the ‘anonymous’ moniker are suspected of working with, or on the behest of, states to launch cyber-attacks.⁶⁸⁷ Such actions engage the rules on state responsibility. In very much the same way as comparable physical actions, wrongful acts committed in cyberspace by states will be subject to these rules.⁶⁸⁸ In addition to this, the ‘effective control’ test set out by the ICJ in the *Nicaragua* case will apply in cases regarding the use of non-state groups to launch such attacks.⁶⁸⁹ This can be straightforwardly transposed from the rules which have developed in relation to kinetic actions. The difficulties, as already discussed above, are in relation to the high levels of deniability that cyber-attacks afford the aggressor.

Professor Tsagourias suggests the lengths required to go to for the attribution necessary in relation to cyber-attacks, emphasising technical, political, and legal aspects, needed to be met.⁶⁹⁰ With only an indifferent reference to it, he appears to overlook a crucial issue in this area – time. When a state is dealing with a cyber-attack in real time, it is likely to be unprepared. During the resulting confusion, the time acquiring the technical, political, and legal requirements of attribution that he suggests, pales into insignificance. Hypothetically speaking, if a state is suffering a prolonged cyber-attack to its critical national infrastructure, whereby the only way to stop it would be to act in self-defence, the luxury of thorough and in depth investigation is one which the victim state could simply not afford. Such actions in self-defence,

available online at, <<https://www.theguardian.com/technology/2015/nov/17/anonymous-war-isis-hacktivist-group-confirms>> accessed 15 April 2016.

⁶⁸⁷ Although this cannot be conclusively confirmed due to attributional issues, attacks such as ‘Turla’ and ‘DarkSeoul’ have been linked to states working through non-state actors. See, Elizabeth Hanford, ‘The Cold War of Cyber Espionage’ (2014) 20(1) Public Interest Law Reporter 22, p. 22. See also, Daniel Nguyen, ‘State Sponsored Cyber Hacking and Espionage’ (2015) ICTN 4040.

⁶⁸⁸ International Law Commission, ‘Articles on Responsibility of states for Internationally Wrongful Acts’ (2001) A/56/10, Vol.I/ch.IV.

⁶⁸⁹ n. 675 (*Nicaragua Case*), p.116-7. *Contra* the ‘overall control test’ of the ICTY in *Prosecutor v Dusko Tadić*, Appeals Chamber, Judgment, 15 July 1999 (Case no. IT-94-1-A) (Hereinafter *Tadić Case*), para. 131. The ICJ reaffirmed the ‘effective control’ test in *Application of the Convention on the Prevention and Punishment of the Crime of Genocide (Bosnia and Herzegovina v Serbia and Montenegro)*, 26 February 2007, ICJ Rep, p. 43 (Hereinafter *Bosnia Genocide Case*), para. 406.

⁶⁹⁰ n. 655 (Tsagourias, 2012).

without true attributional confirmation, is of concern, however.⁶⁹¹ As such, a general test for attribution based on a principle such as on the ‘balance of probabilities’ could be a solution. If this was strictly applied to acting in self-defence only through cyberspace, such a test would likely be more effective than if it allowed for a reaction in the kinetic world. It does face obstacles, however, in that due to the nature of the sphere in which cyber-attacks occur, it is not difficult to frame an attack as coming from a certain state or actor. This does not mean that the test on the ‘balance of probabilities’ could never operate successfully, as contextual and sociological factors can also be applied.

If a conflict is perceived to exist following a cyber-attack reaching the threshold for an armed attack, the LOAC will become relevant. The Martens Clause is intended to deal with areas of the LOAC that are not yet regulated and allows for the principles already in place to be applied.⁶⁹² Indeed, there is a distinct lack of new legal frameworks or legislation regarding cyberspace under international law.⁶⁹³ This has, in turn, had the inevitable consequence of international lawyers feeling as though cyberspace is an area left “dangerously unregulated.”⁶⁹⁴ We will now look at the regulations that can be applied to cyberspace in the event of an armed conflict.

⁶⁹¹ Evidential difficulties such as this were recognised by the ICJ in, *Corfu Channel Case (United Kingdom of Great Britain and Northern Ireland v Albania)*, Merits, 9 April 1949, ICJ Rep, p. 4, p. 18.

⁶⁹² Modern version of the Martens Clause is embodied in, Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (adopted 8 June 1977, entered into force 7 December 1978) 1125 UNTS 3 (Hereinafter Additional Protocol I), Article 1(2). The Martens Clause first appeared in, Convention (II) with Respect to the Laws and Customs of War on Land and its annex: Regulations concerning the Laws and Customs of War on Land (adopted 29 July 1899, entered into force 4 September 1900) 187 CTS 429 (Hereinafter Hague Convention II 1899), Preamble.

⁶⁹³ Jean d’Aspremont, ‘Cyber Operations and International Law: An Interventionist Legal Thought’ (2016) 21 *Journal of Conflict and Security Law* (forthcoming) ACIL Research Paper 2016-11, <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2765408> accessed 25 April 2016, p. 3.

⁶⁹⁴ *ibid.*

At a basic level, an armed conflict exists when a group acts towards a common goal of injuring, killing, or destroying an opposition.⁶⁹⁵ If a cyber-based attack is intended to cause any of those end results in the physical world, it is well-settled that it should be considered a use of force. The problems, however, include how the principle of distinction can be retained and attributional issues pertaining to cyber-attacks. To launch such an attack is not necessarily an intrinsically military capability and this, therefore, raises questions relating to civilians directly participating in hostilities (DPH),⁶⁹⁶ through their actions in cyberspace. This further adds to the conglomeration of issues that cyber warfare has brought to international law.

Prior to discussion of the rules pertaining to the LOAC and its potential application to cyber warfare, it should first be considered whether a cyber weapon can be treated as a weapon at all. Article 36 of Additional Protocol I,⁶⁹⁷ addresses new weaponry. This includes the duty to determine whether it would be prohibited in any way by Additional Protocol I. As has been established above, cyber-attacks can occur across a range of methods against an array of targets. If a cyber-attack was aimed at the release of a prohibited weapon under international law, such as incendiary weapons,⁶⁹⁸ or could have an effect on a work or installation containing dangerous forces,⁶⁹⁹ then the legality of the cyber weapon itself needs to be considered, not just the use of it, as the literature has focused on thus far.⁷⁰⁰

There have been attempts to create a document that can be applied to cyberwarfare. One such example is the Tallinn Manual of the International Law Applicable to Cyber Warfare (the Tallinn Manual).⁷⁰¹ The Manual argues for the application of the LOAC to

⁶⁹⁵ n. 670 (Schmitt, 2002), p. 373.

⁶⁹⁶ n. 692 (Additional Protocol I), Article 51(3).

⁶⁹⁷ *ibid* (Additional Protocol I), Article 36.

⁶⁹⁸ Jean-Marie Henckaerts, Louise Doswald-Beck, *ICRC: Customary International Humanitarian Law, Volume I: Rules* (Cambridge University Press: Cambridge, 2005) (Hereinafter ICRC Study on CIHL), Rule 84.

⁶⁹⁹ n. 692 (Additional Protocol I), Article 57.

⁷⁰⁰ Katharina Ziolkowski, 'Computer Network operations and the law of armed conflict' (2010) 49 *Military Law and Law of War Review* 47, p. 69.

⁷⁰¹ n. 675 (Tallinn Manual).

cyber-attacks and the principles that apply in physical warfare to be applied also to cyber warfare. Applying the rules in place for physical conflict, to the fifth domain of warfare, would appear straightforward. However, this overlooks the complexities within the LOAC itself. Much of the LOAC is contested. When adding to it the novel issues that cyberwarfare creates, it does not become any more straightforward. This is complicated further, as the Tallinn manual's authors come from NATO Member States, and despite declarations to the contrary, can be seen as representing the attitudes and values of NATO itself.⁷⁰² We will now look at some of the ways in which cyber-attacks can create issues within the LOAC.

One of the cornerstones of the LOAC is the principle of distinction.⁷⁰³ This principle is important as it relates to who can be targeted in an armed conflict. It protects civilians in prohibiting the intended and deliberate targeting of them.⁷⁰⁴ Cyberspace is accessible by all and does not face the same physical restraints as kinetic attacks. This may have numerous military advantages, including a reduction of personnel on the physical battlefield, and the ability to launch attacks from one part of the world to have an effect in another. There is, however, the danger of attacks spilling into the civilian sphere, and also those which could release dangerous forces.⁷⁰⁵ Cyberspace is inherently of a dual-use character. An attack aimed at a military target, therefore, could plausibly have residual effects in the civilian cyberspace domain. Attacks against Supervisory Control and Data Acquisition (SCADA) systems have been known to cause disruption and have the potential for disastrous consequences.⁷⁰⁶ This can also occur transnationally over a borderless online world. An attack in

⁷⁰² David P Fidler, 'NATO, Cyber Defense, and International Law' (2013) 4 St. John's Journal of International & Comparative Law 1.

⁷⁰³ n. 677 (*Nuclear Weapons* Advisory Opinion), p. 257, para. 78. See also, n. 692 (Additional Protocol I), Article 48. See also n. 698 (ICRC Study on CIHL), Rule 1.

⁷⁰⁴ n. 692 (Additional Protocol I), Article 51(2).

⁷⁰⁵ For example, see, *ibid* (Additional Protocol I), Articles 52, 54 and 56. The aforementioned Stuxnet attack could have potentially led to the release of 'dangerous forces', for example.

⁷⁰⁶ n. 671 (Andress and Winterfeld, 2013), p. 58-9. See also Calvery Bowers, 'A Plan for SCADA Security to deter DoS Attacks' (2005) in Calvery Bowers et al, s.1 Proceedings of the Department of Homeland Security, R&D Partnering Conference.

cyberspace has the potential of becoming indiscriminate and as such violating one of the main tenets of the LOAC.⁷⁰⁷

The law applicable to cyber warfare faces further challenges when principles such as proportionality are considered. When examining the *jus in bello* (how armed conflicts are to be conducted), the way in which a cyber weapon is used within the conflict needs to be regulated to cover the extent of its use and collateral damage. As has been seen throughout this essay, cyber-attacks can have exponential and potentially uncontrollable effects once released from human control. This has led to some unhelpful suggestions. Professor Schmitt, for example, argues that once a cyber weapon has left the control of the individual operating it, they no longer have control over its effects. As such, any indiscriminate effects occurring thereon because of the cyber weapon is not a result of the weapons system itself and cannot be attributed to the human actor, or who they work on behalf of.⁷⁰⁸ Such an argument appears to be based on a willingness to relinquish all human responsibility for a cyber-attack. This is a dangerous suggestion. It is analogous to blindly dropping many bombs at a military target on a windy day, and once the bombs have been released, the effects they have, be indiscriminate or not, are not the responsibility of those who dropped them. An eccentric hypothetical situation such as this highlights the absurdity of Professor Schmitt's submission.

⁷⁰⁷ n. 692 (Additional Protocol I), Article 51(4). See also, Kelly O'Connell, 'Internet Law: Russian Company Outed as Mother of All Cyber-Crime' (2007) Internet Business Law Services, available online at, https://www.ibls.com/internet_law_news_portal_view.aspx?s=latestnews&id=1887 accessed 7 April 2016.

⁷⁰⁸ n. 670 (Schmitt, 2002), p. 389.

III. Conclusion

The future of cyberspace is both predictable and unpredictable. It is both rosy and gloomy. The potential benefits seem unconstrained and perpetual. Undercutting this is the growing threat to the safety and security of those who operate within cyberspace. As this essay has discussed, there are many actors participating in cyberspace. Rising levels of reports of ‘ransomware’ affecting cyber systems,⁷⁰⁹ and the impending threat of a large-scale attack on portable ‘smart’ phones,⁷¹⁰ ensure an uncertain future for the widespread stability of the domain. From a militaristic perspective, it is now clear that cyberspace is certainly a field from which attacks can originate. It can be used to great effect in individual and coordinated offensive military strikes. More harrowing is the prospect of terrorist organisations acquiring the competence to launch cyber-attacks against high-profile targets and even critical national infrastructure. Such occurrences could have monumental impacts on both civilian and military applications for cyberspace and the domain as a whole. Any attack of this kind would threaten the two most highly venerated tenets within cyberspace – online freedoms and online sovereignty. Attributional issues have been highlighted as a critical obstacle to the further development of legal regulation in cyberspace. This essay has demonstrated some of the inadequacies of the current legal framework pertaining to cyberspace, and offered suggestions of improvements that can be made in both the civilian sphere and on the military stage. The not-too-distant future will be sure to illuminate further developments in both the capabilities of the cyber world and the principles used to regulate it.

⁷⁰⁹ Alexandre Gazet, ‘Comparative analysis of various ransomware virii’ (2010) 6(1) *Journal in Computer Virology* 77. See also, Seung Lee, ‘Your money or your files: the growing threat of ransomware’ *The Independent* (London, 19 May 2016) available online at, <<http://www.independent.co.uk/life-style/gadgets-and-tech/features/your-money-or-your-files-the-growing-threat-of-ransomware-a7035421.html>> accessed 20 May 2016.

⁷¹⁰ Stuart Dredge, ‘How could you become a victim of cybercrime in 2015’ *The Guardian* (London 24 December 2014), available online at, <<https://www.theguardian.com/technology/2014/dec/24/cybercrime-2015-cybersecurity-ransomware-cyberwar>> accessed 7 April 2016.

A critical assessment of the separability doctrine, its impact, and application

*Fiona Winnifred Lakareber*⁷¹¹

The separability doctrine has, in recent years, gained significant popularity in international commercial arbitration, and has been credited for the general increase in arbitration cases. There are nevertheless a number of drawbacks associated with the application and impact of the doctrine, prompting critics to advocate its abolition. The main challenges are both historical and contemporary, and include a fragmented approach by courts to separability issues and the absence of clear guidance on how to ascertain consent by the parties to the arbitration clause. Using examples from the United Kingdom and the United States, this paper evaluates some of these drawbacks and takes the view that separability should not be abolished because its benefits outshine the existing challenges by a considerable magnitude. It is for this reason that states have steadily moved away from being indifferent, and sometimes hostile to arbitration to providing legal tools for the enforcement of the separability doctrine. Moreover, even the most pronounced critics of separability acknowledge the immense difficulty the international business community would be faced with were the doctrine to be abolished. This paper, therefore, urges critics to join hands with pro-separability scholars, courts and states, to design effective, long-lasting solutions to the loopholes pertaining to the application and impact of the doctrine. Parties to commercial contracts are equally called upon to clearly and concisely articulate their intention to arbitrate in order to enhance the efficacy of separability and, ultimately, their choice of arbitration to resolve ensuing disputes.

⁷¹¹ LL.M. Candidate, The University of Manchester, School of Law.

I. Introduction

The separability doctrine (“separability”) is one “of most widely recognised principles in international commercial arbitration.”⁷¹² Separability means the arbitration clause of a contract is severable from the main contract and remains in force despite the invalidity, nullity or termination of the main contract.⁷¹³ It is broadly recognised in international conventions, institutional arbitration rules, and the national laws of several developed and developing states.⁷¹⁴ By protecting the arbitration clause from the defects of the main contract, separability ensures that the clause only fails on grounds which directly affect it.⁷¹⁵ This is important because one of the situations in which an arbitration clause plays a crucial role is when the validity of the main contract is disputed.⁷¹⁶ For this reason, Lord MacMillan has stated that a contested contract “survives for the purpose of measuring the claims arising out of the breach, and the arbitration clause survives for determining the mode of their settlement.”⁷¹⁷

Despite the wide acceptance of separability across developed states, several challenges have arisen in its application, resulting in inconsistent and confusing court decisions.⁷¹⁸ The distinction that states drew between void and voidable contracts was historically one of the main challenges.⁷¹⁹ Although in the United States, *Buckeye Check Cashing Inc. v Cardegna*

⁷¹² Ozlem Sulser, ‘The Jurisdiction of the Arbitral Tribunal: A Transnational Analysis of the Negative Effect of Competence-Competence’ [2009] *Macquarie Journal of Business Law* 119, 119.

⁷¹³ Thomas D. Grant, ‘International Arbitration and English Courts’ [2007] 56 (04) *International and Comparative Law Quarterly* 871, 873.

⁷¹⁴ Alan Redfern and Martin Hunter, *Redfern and Hunter on International Arbitration* (6th edn, Oxford University Press 2015) 106.

⁷¹⁵ *Fili Shipping Co Ltd and Others v Premium Nafta Products Ltd – On Appeal from Fiona Trust and Holding Corp’n and Others v Privalov and others (‘Fiona’)* [2007] Bus LR 1719 [35] (Lord Hope of Craighead).

⁷¹⁶ Redfern and Hunter (n 714) 104.

⁷¹⁷ *Heyman and Another v Darwins, Limited* [1941] AC 356, 374 (Lord Macmillan).

⁷¹⁸ Richard C. Reuben, ‘First Options, Consent To Arbitration, and the Demise of Separability: Restoring Access to Justice for Contracts with Arbitration Provisions’ [2003] 56 *SMUL Rev* 819, 829.

⁷¹⁹ Carl Svernlöv, ‘What Isn’t, Ain’t: The Current Status of the Doctrine of Separability’ [1991] 8 *J Int’l Arb* 37, 45.

(‘*Buckeye*’)⁷²⁰ abolished the distinction that *Prima Paint v Flood & Conklin* (‘*Prima*’)⁷²¹ made between void and voidable contracts, certain questions remain. Indeed, *Buckeye* itself, and other landmark cases such as *Fili Shipping Co Ltd and Others v Premium Nafta Products Ltd – On Appeal from Fiona Trust and Holding Corpn and Others v Privalov and others* (‘*Fiona*’),⁷²² introduced additional problems. Some of the historical and new challenges include piecemeal determination of issues by courts,⁷²³ restrictions relating to contract formation disputes, and lack of clear guidance on ascertaining specific assent by the parties to the arbitration clause.⁷²⁴

These gaps have led scholars like Ware to call for the repeal of separability.⁷²⁵ Moreover, the benefits that flow from separability (if upheld) are themselves tainted. Against this background, this paper seeks to evaluate the unresolved issues pertaining to the application and impact of separability. It also argues for the retention of separability despite its challenges, by linking the doctrine to the general increase in the popularity of arbitration.⁷²⁶ This linkage is currently lacking in the literature, which therefore underestimates the significance of separability to the growth in arbitration.

The United Kingdom is a global centre for arbitration.⁷²⁷ It has broadly embraced separability in a bid to end the more than 398-year debate on arbitration clauses.⁷²⁸ Conversely, the US has

⁷²⁰ [2006] 546 US 440.

⁷²¹ [1967] 388 US 395.

⁷²² *Fiona* (n 715).

⁷²³ *Buckeye* (n 720) footnote 1.

⁷²⁴ Alan Scott Rau, ‘Separability in the United States Supreme Court’ [2006] Stockholm International Arbitration Review (Forthcoming) 17 <<http://ssrn.com/abstract=893601>> accessed 16 March 2016.

⁷²⁵ Stephen J. Ware, ‘Arbitration Law’s Separability Doctrine After *Buckeye Check Cashing Inc. v Cardegna*’ [2007] 8 Nevada Law Journal 107, 119.

⁷²⁶ Gary Born and Wendy Miles, ‘Global Trends in International Arbitration’ (2016), <http://www.wilmerhale.com/uploadedFiles/WilmerHale_Shared_Content/Files/Editorial/Publication/GlobalTrends_InternationalArbitration.pdf> accessed 14 April 2016. See also LCIA Registrar’s Report 2015, 1.

⁷²⁷ Lucy Scott-Moncrieff, ‘Arbitration: What Makes London So Special?’ (*The Law Society*, 15 November 2012) <<https://www.lawsociety.org.uk/news/speeches/arbitration-what-makes-london-so-special/>> accessed 08 April 2016.

⁷²⁸ Adam Samuel, ‘Agora: Thoughts on *Fiona Trust*’ [2008] 24(3) *Arbitration International* 489, 489.

applied separability rather restrictively.⁷²⁹ On this basis, coupled with their shared common law traditions, the UK and the US will be used to illustrate the arguments presented below.

For the avoidance of doubt, this paper does not aim to undertake a comparative analysis of the UK and the US' treatment of separability. It does not purport to provide a checklist of all existing challenges with the application and impact of separability. Rather, it provides an evaluation of the main difficulties that the general international business community is faced with, supported with examples from the UK and the US. The practical challenges pertaining to special categories of contracts where one of the parties has a weaker bargaining power than the other are therefore outside the scope of the paper. These include consumer, insurance and employment contracts into which arbitration clauses are unilaterally imposed so that the counterparty's ability to seek recourse to the courts is reduced or eliminated.

This paper has four other sections. Section II provides an overview of separability and analyses its rationale. This background information is important for understanding separability and for evaluating, ultimately, whether the practical gaps in its application render it worthless. A critique of the key legal provisions on separability is essential for understanding the theoretical position of the law, and assessing the extent to which they promote the doctrine. This is also done in Section II, which analyses the UK's Arbitration Act 1996, the US's Federal Arbitration Act 1925, and the UNCITRAL Model Law on International Commercial Arbitration 1985.⁷³⁰ The section further analyses these laws regarding any loopholes that contribute to the difficulties faced in applying separability. Section III assesses the impact of separability (if upheld) and comments on the extent to which this accords with its rationale. The application of separability is then assessed in Section IV with reference to landmark decisions including *Buckeye*, *Prima*, and *Fiona*. The

⁷²⁹ Tanya J. Monestier, "Nothing Comes of Nothing" . . . Or does it??? A Critical Re-Examination of the Doctrine of Separability in American Arbitration' [2001] 12 Am Rev Int'l Arb 223, 223.

⁷³⁰ Arbitration Act 1996; Federal Arbitration Act 1925; UNCITRAL Model Law on International Commercial Arbitration 1985.

purpose of the Section is to evaluate the questions that these cases leave unresolved, or raise, thereby denying parties the opportunity to reap all the benefits that separability would otherwise offer. Since there is little in the literature currently linking separability to the popularity of arbitration, Section IV synthesises the available evidence and evaluates whether the challenges in the application of the doctrine overshadow its benefits. Section V concludes by summarising the main points presented throughout the paper, and makes some recommendations for resolving the present challenges.

II. Overview, Rationale and Legal Recognition of Separability

(i) Overview and Rationale

Historically, states demonstrated a significant mistrust of arbitration,⁷³¹ viewing it as a forum that eliminated national courts' authority to decide disputes.⁷³² As such, it was common for states to be either indifferent or hostile towards arbitration.⁷³³ The agreement to arbitrate was consequently treated as unenforceable⁷³⁴ and could be revoked by either party whenever they wished.⁷³⁵ Although courts later began referring complex issues to arbitrators, they continued to intervene directly in tribunals' proceedings, making such referrals merely an extension of the judicial process.⁷³⁶

⁷³¹ Pierre H. Bergeron, 'At the Crossroads of Federalism and Arbitration: The Application of Prima Paint to Purportedly Void Contracts' [2004] 93 Ky LJ 423, 423.

⁷³² Piero Bernardini, 'The Arbitration Clause of an International Contract' [1992] 9 (2) J Int'l Arb 45, 45.

⁷³³ Phillip Landolt, 'The Inconvenience of Principle: Separability and Kompetenz-Kompetenz' [2013] 30 (5) J Int Arb 511, 512.

⁷³⁴ Lord Mustill MJ, 'Arbitration: History and Background' [1989] 6 (2) J Int'l Arb 43, 49.

⁷³⁵ Adam Samuel, 'Separability and the US Supreme Court Decision in *Buckeye v. Cardegna*' [2006] 22 (3) Arbitration International 477, 478.

⁷³⁶ Mustill (n 734) 46.

Over time, the complementary role of arbitration in dispute resolution, particularly its ability to reduce congestion in courts, was recognised.⁷³⁷ This led to the emergence of separability to uphold parties' desires⁷³⁸ to have their disputes resolved in one forum,⁷³⁹ and to make arbitration an effective dispute resolution mechanism by minimising the intervention of courts.⁷⁴⁰ Separability is the state's way of lending support to disputing parties by facilitating enforcement of their agreement to avoid the public justice system.⁷⁴¹ This protects the parties from the uncertainties of national procedural laws and requires courts to interpret broadly arbitration clauses.⁷⁴² So in *Fiona*, the UK Supreme Court held that courts should focus on fulfilling parties' reasonable commercial expectations to resolve all their disputes in a single forum, and not on analysing the meaning of words used in the arbitration clause.⁷⁴³ Although the court expressed this view not to advocate separability *per se*, its decision is ultimately pro-separability.⁷⁴⁴

The 'reasonable commercial expectations' criterion empowers courts to imply consent of the parties to arbitrate where there is no clear expression of such intent.⁷⁴⁵ Whether such an inference always represents the true will of parties remains doubtful.⁷⁴⁶ The consequence of this uncertainty is the line of

⁷³⁷ Jill I. Gross, 'Justice Scalia's Hat Trick and the Supreme Court's Flawed Understanding of Twenty-First Century Arbitration' [2015] 81 (1) Brooklyn Law Review 111, 117.

⁷³⁸ Sulser (n 712) 124.

⁷³⁹ Alan Scott Rau, 'Everything You Really Need to Know about "Separability" in Seventeen Simple Propositions' [2003] 14 Am Rev Int'l Arb 182.

⁷⁴⁰ Stephen M. Schwebel, *International Arbitration: Three Salient Problems* (Grotius Publications Limited 1987) 4.

⁷⁴¹ William W. Park, 'The Arbitrability Dicta in First Options v. Kaplan: What Sort of Kompetenz-Kompetenz Has Crossed the Atlantic?' [1996] 12 (2) Arbitration Journal 137, 139.

⁷⁴² Philippe Le Boulanger, 'The Arbitration Agreement: Still Autonomous' in Albert Jan van den Berg (ed), *International Arbitration 2006: Back to Basics? ICCA Congress in Series 2006 Montreal 13* (Kluwer Law International 2007) 3.

⁷⁴³ *Fiona* (n 715) [27] (Lord Hope of Craighead).

⁷⁴⁴ Samuel, 'Agora: Thoughts on Fiona Trust' (n 728) 493.

⁷⁴⁵ Reuben (n 718) 846.

⁷⁴⁶ Pierre Mayer, 'The Limits of Severability of the Arbitration Clause' in Albert Jan van den Berg (ed), *Improving the Efficiency of Arbitration Agreements and Awards: 40 Years of Application of the New York Convention* (Kluwer Law International 1999) 261, 264.

reasoning that, if parties wanted two agreements, they would prepare and sign two different documents.⁷⁴⁷ But had they done so, there would be no need for separability since the arbitration agreement would already be separate from the main contract.

More specifically, Reuben contends that *Prima* violated Congress' requirement for actual consent to arbitration agreements, by introducing implied consent.⁷⁴⁸ Although this argument is reasonable, it is also possible in practice that parties do know and consent to the arbitration clause when they sign the main contract. Knowledge is a crucial element of consent and thus it is not worthwhile to demand full consent before concluding that such clauses should be upheld.⁷⁴⁹

Separability also guards against the delaying tactics a party may employ by alleging invalidity of the main contract containing the arbitration clause.⁷⁵⁰ This constitutes abuse of process and intolerable behavior,⁷⁵¹ which would otherwise result in the early involvement of courts in disputes that parties agreed to resolve by arbitration.⁷⁵²

(ii) *Recognition under the Federal Arbitration Act 1925 ('US-FAA')*

In the US, separability is not expressly provided for under the US-FAA⁷⁵³ but is derived from sections 2 and 4 of the Act.⁷⁵⁴ The US-FAA equates arbitration agreements to ordinary contracts which may be rendered invalid, revocable or unenforceable on the same

⁷⁴⁷ *ibid* 261.

⁷⁴⁸ Reuben (n 718) 844.

⁷⁴⁹ Brian H. Bix, 'Contracts' in Franklin G. Miller and Alan Wertheimer (eds) *The Ethics of Consent: Theory and Practice* (Oxford University Press 2010) 251, 253 – 255.

⁷⁵⁰ Schwebel (n 740) 4.

⁷⁵¹ Svernlöv (n 719) 37.

⁷⁵² Sulser (n 712) 128.

⁷⁵³ Jack Graves and Yelena Davydan, 'Competence-Competence and Separability-American Style' in Stefan Michael Kroll, Loukas A. Mistelis, Pilar Perales Viscasillas and Vikki Rogers (eds), *International Arbitration and International Commercial Law: Synergy, Convergence and Evolution - Chapter 8*. (Wolters Kluwer Law & Business/Aspen Publishers 2011) 157, 158.

⁷⁵⁴ Federal Arbitration Act 1925, s.2 and s.4.

legal or equitable grounds as any other contract.⁷⁵⁵ Further, a party may apply for an order compelling arbitration which courts are bound to honour unless there is a failure, neglect, or refusal to perform the arbitration agreement, or the construction of the arbitration agreement is in issue.⁷⁵⁶

It should be noted that the US-FAA applies both to state courts and federal courts.⁷⁵⁷ Unfortunately, some state laws may contradict the US-FAA on separability,⁷⁵⁸ because the US-FAA prevails over inconsistent state laws,⁷⁵⁹ states are burdened in terms of time and cost to align their arbitration laws with the statute. Further, there is no federal common law of contract in the US, yet the US-FAA prescribes contract law as the basis for determining the status of arbitration agreements.⁷⁶⁰ Contract law questions are thus answered in accordance with individual states' laws.⁷⁶¹ Whilst states must abide by the pro-arbitration policy of the US-FAA and most have enacted separability laws, inconsistent application of separability and uncertain outcomes across states are still highly likely.⁷⁶²

(iii) Recognition under the Arbitration Act 1996 ('UK-AA')

Section 7 of the UK-AA is the foundation of separability in the UK.⁷⁶³ It recognises that an arbitration agreement is a distinct agreement and is not affected by the invalidity, non-existence, or ineffectiveness of the contract which it is, or is intended to be a part of.⁷⁶⁴

However, Section 7 is silent on whether the reference to 'non-existence' of the main contract means the contract never

⁷⁵⁵ *ibid* s 2.

⁷⁵⁶ *ibid* s 4.

⁷⁵⁷ *Buckeye* (n 720) [B].

⁷⁵⁸ Samuel, 'Separability and the US Supreme Court...' (n 735735735) 480.

⁷⁵⁹ *Nitro-Lift Technologies, L. L. C. v Eddie Lee Howard Et Al*, [2012] 133 S.Ct. 500, 504.

⁷⁶⁰ Federal Arbitration Act 1925, s 2.

⁷⁶¹ Park, 'The Arbitrability Dicta in *First Options v. Kaplan*...' (n 741) 141.

⁷⁶² Reuben (n 718) 852.

⁷⁶³ Grant (n 713) 873.

⁷⁶⁴ Arbitration Act 1996, s.7.

came into existence or that it was properly formed but simply ceased to exist. This has resulted in the narrow view that separability should only apply where the main contract is properly formed but subsequently ceases to exist,⁷⁶⁵ for example, due to lapse of time.⁷⁶⁶ Fortunately, the UK Supreme Court, recognising that Parliament did not intend to impose unnecessary limitations on the parties' desire to arbitrate disputes, took the broader view that arbitrators should decide all questions about the status of the main contract.⁷⁶⁷

The UK-AA is heavily rooted in the will of the parties to make decisions freely regarding arbitration matters.⁷⁶⁸ Consequently, parties may disregard separability if they wish.⁷⁶⁹ On the one hand, this right to decide against separability is justifiable given that the foundation of arbitration is freedom of contract.⁷⁷⁰ On the other hand, the public good associated with arbitration - easing courts' caseload - would be lost if the parties to a defective contract elected not to honour separability.⁷⁷¹

(iv) Recognition under the UNCITRAL Model Law on International Commercial Arbitration 1985 ('Model Law')

The Model Law aims to promote uniform application of arbitration procedures across states so as to satisfy the requirements of international arbitration practice.⁷⁷² It recognises that an arbitration clause contained in a contract is a distinct agreement from the other terms of the contract and is therefore not rendered invalid merely because the main contract is null and void.⁷⁷³

⁷⁶⁵ Sulser (n 712) 123.

⁷⁶⁶ Chris Turner, *Unlocking Contract Law* (4th Revised edn, Taylor and Francis 2014) 350.

⁷⁶⁷ *Fiona* (n 715) [10] (Lord Hoffman).

⁷⁶⁸ Charles Chatterjee, 'The Reality of Party Autonomy Rule in International Arbitration' [2003] 20 (6) *Journal of International Arbitration* 539, 542.

⁷⁶⁹ Arbitration Act 1996, s 7.

⁷⁷⁰ Thomas E. Carbonneau, 'The Exercise of Contract Freedom in the Making of Arbitration Agreements' [2003] 36 *Vand J Transnat'l L* 1189, 1195.

⁷⁷¹ Gross (n 737) 117.

⁷⁷² UNCITRAL Model Law on International Commercial Arbitration 1985, Resolutions adopted by the General Assembly, p. vii.

⁷⁷³ *ibid*, art 16 (1).

It is interesting to note that the Model Law discusses separability under the title ‘competence of arbitral tribunal to rule on its jurisdiction’.⁷⁷⁴ This title relates to the doctrine of competence-competence (discussed in detail in Section III) and may result in the incorrect inference that competence-competence and separability are the same.⁷⁷⁵ Whereas separability splits the arbitration clause from the main contract, competence-competence allows the tribunal to proceed with hearing the dispute even if its power to do so is challenged by a party.⁷⁷⁶ Nevertheless, the doctrines share the common goal⁷⁷⁷ of avoiding abuse of process by the parties through delaying tactics.⁷⁷⁸

Finally, unlike the UK-AA, the Model Law does not allow parties to disregard separability.⁷⁷⁹ In effect, the Model Law goes further than the UK-AA towards ensuring that separability is upheld in all instances.⁷⁸⁰

(v) Improving the Law on Separability

Arbitration clauses exclude the powers of state courts to deal with parties’ disputes,⁷⁸¹ at least initially.⁷⁸² Also, the arbitration clause constitutes a distinct agreement from the main contract.⁷⁸³ However, neither of these factors characterises the other terms of the main contract. Consequently, an arbitration clause cannot be equated to an ordinary contract term to which the rules of contractual interpretation apply. The assertion that the arbitration clause “is

⁷⁷⁴ *ibid.*

⁷⁷⁵ John J. Barcelo III, ‘Who Decides the Arbitrators’ Jurisdiction? Separability and Competence-Competence in Transnational Perspective’ [2003] 36 Vand. J Transnat’l L 1115, 1116.

⁷⁷⁶ William W. Park, ‘The Arbitrator’s Jurisdiction to Determine Jurisdiction, I. Introduction: The Limits of Language’ in Albert Jan van den Berg (ed), *International Arbitration 2006: Back to Basics?* ICCA Congress Series, Volume 13 (Kluwer Law International 2007) 56.

⁷⁷⁷ Barcelo III (n 775) 1116.

⁷⁷⁸ Svernlöv (n 719) 37.

⁷⁷⁹ Landolt (n 733) 523.

⁷⁸⁰ *ibid* 524.

⁷⁸¹ Bernardini (n 732) 46.

⁷⁸² Barcelo III (n 775) 1118.

⁷⁸³ Leboulanger (n 742) 1.

merely one of several clauses” in the main contract, therefore, undermines their importance.⁷⁸⁴

As such, the special nature of the arbitration clause warrants distinct rules for ascertaining the parties’ true intentions. These rules could be included among the useful aspects of separability that Landolt proposes should be enacted into law.⁷⁸⁵ A prerequisite for this, however, is the identification of coherent principles which, as Section IV shows, has not yet been done.

In addition to arbitration clauses, the US-FAA⁷⁸⁶ and the UK-AA⁷⁸⁷ provide for submission agreements. A submission agreement is one “by which the parties to an existing dispute submit it to arbitration.”⁷⁸⁸ They are not part of the main contract, raising the question whether separability should apply to them at all. Referring to Section 7 of the UK-AA, Sulser, for example, states that separability does not make a distinction between arbitration clauses and submission agreements.⁷⁸⁹ This conflicts with the rather restrictive definition she adopts for her paper, to the effect that separability means “...the arbitration clause/agreement in a contract that is deemed to be separate from the main contract...”⁷⁹⁰ The words ‘in a contract’ can only mean an arbitration clause. Such inconsistencies are an indication of the lack of clarity on whether separability should apply both to arbitration clauses and submission agreements.

The application of separability also presents practical problems in respect of arbitration clauses but not submission agreements.⁷⁹¹ This exemplifies how separability is redundant

⁷⁸⁴ Mayer (n 746) 261.

⁷⁸⁵ Landolt (n 733) 512.

⁷⁸⁶ Federal Arbitration Act 1925, s.2.

⁷⁸⁷ Arbitration Act 1996, s.7.

⁷⁸⁸ United Nations Conference on Trade and Development (UNCTAD): Dispute Settlement - International Commercial Arbitration; 5.2 The Arbitration Agreement (Ref: UNCTAD/EDM/Misc.232/Add.39) <http://unctad.org/en/Docs/edmmisc232add39_en.pdf> accessed 30 March 2016.

⁷⁸⁹ Sulser (n 712) 123.

⁷⁹⁰ *ibid* 122.

⁷⁹¹ Svernlöv (n 719) 38.

when it comes to submission agreements. The most reasonable explanation is that submission agreements are already separate from the main contract, with their own forms and signatures, and parties are likely to have deliberated on their implications in much more detail than with arbitration clauses.

Besides, the celebrated separability cases [*Prima, Buckeye, Fiona and Harbour Assurance Co (UK) Ltd v Kansa General International Insurance Co Ltd* ('*Harbour*')⁷⁹²] all relate to arbitration clauses. *Harbour*, for instance, concerned illegality and the court cited the need to honour the parties' choice of arbitration and to ensure a one-stop forum for the resolution of their dispute as 'powerful commercial reasons' for upholding the arbitration clause, despite the illegality of the main contract.⁷⁹³

In this vein, it is plausible to suggest that separability should only apply to arbitration clauses, which are embedded in the main contract and are capable of being 'separated' from it. The Model Law follows this view by confining its separability provision to the arbitration clause.⁷⁹⁴ After all, "the gist of the [separability] doctrine is that an arbitration clause [is contained] in a larger contract [and] must be carved out, severed from the larger contract, and examined separately".⁷⁹⁵

⁷⁹² [1993] QB 701.

⁷⁹³ *ibid* 724 (Hoffman LJ).

⁷⁹⁴ UNCITRAL Model Law on International Commercial Arbitration 1985, s.16 (1).

⁷⁹⁵ *Schumacher Homes Of Circleville Inc. v John Spencer And Carolyn Spencer* [2016] No. 14-0441, W Va June 13, 2016 (Chief Justice Ketchum).

III. Impact of Separability

(i) Overview

The primary impact of separability is the detachment of the arbitration clause from the main contract.⁷⁹⁶ One possible outcome is that the arbitration clause may be invalid despite the main contract being valid.⁷⁹⁷ This arises, for example, where the parties are prohibited from excluding the jurisdiction of state courts by going to arbitration, say on public policy grounds.⁷⁹⁸ In the alternative, the fate of the arbitration clause may follow that of the main contract but only if the factors affecting the main contract specifically touch upon the arbitration clause.⁷⁹⁹ The third, and perhaps best case scenario, is where the arbitration clause remains valid despite the invalidity of the main contract.⁸⁰⁰ This protects the arbitration clause from expiring or being terminated along with the main contract.⁸⁰¹ In such cases, the consequences that flow from separability include competence-competence and the choice of law governing the arbitration clause.⁸⁰²

(ii) Competence-competence

Competence-competence is a consequence⁸⁰³ and procedural tool⁸⁰⁴ of the separability doctrine but it has developed much slower than separability.⁸⁰⁵ This is because competence-competence directly takes away state courts' jurisdiction.⁸⁰⁶ It is therefore considered to be more controversial than separability and has led to states adopting and applying it inconsistently.⁸⁰⁷

⁷⁹⁶ Gerald Aksen, 'Prima Paint v. Flood & Conklin-What Does It Mean?' [1968] 43 (1) John's L. Rev. 1, 7.

⁷⁹⁷ Jan Paulsson, *The Idea of Arbitration* (1st edn, Oxford University Press 2013) 62.

⁷⁹⁸ Barcelo III (n 775) 1119.

⁷⁹⁹ Paulsson (n 797) 63.

⁸⁰⁰ *ibid.*

⁸⁰¹ Sulser (n 712) 122.

⁸⁰² Leboulanger (n 742) 7.

⁸⁰³ Reuben (n 718) 828.

⁸⁰⁴ Leboulanger (n 742) 8.

⁸⁰⁵ Graves and Davydan (n 753) 158.

⁸⁰⁶ Bernardini (n 732732) 45.

⁸⁰⁷ Barcelo III (n 775) 1123.

Founded on the rule that every judge is competent to determine his or her own jurisdiction,⁸⁰⁸ competence-competence was “developed to enhance the effectiveness of arbitration as a dispute settlement mechanism.”⁸⁰⁹ Competence-competence empowers the arbitral tribunal to decide on its own jurisdiction in the event that one of the parties to the arbitration challenges it.⁸¹⁰ It also enables arbitrators to determine whether the arbitral tribunal is appropriately constituted⁸¹¹ and to exercise their powers despite the potential defects of the main contract.⁸¹² In short, competence-competence enables tribunals to continue hearing parties’ disputes without having to wait for court rulings on their jurisdiction.⁸¹³

The basis of competence-competence is the parties’ agreement to arbitrate.⁸¹⁴ Separability supports this by severing the arbitration clause from the main contract and enabling arbitrators to apply their powers despite objections to the validity of the main contract.⁸¹⁵ Competence-competence has been so widely recognised⁸¹⁶ that it is provided for in both the UK-AA⁸¹⁷ and Model Law.⁸¹⁸ The US, however, has never formally incorporated it into the US-FAA, leaving it to be implied from the decisions in *Prima* and *First Options of Chicago Inc. v Kaplan* (*‘First Options’*).⁸¹⁹

The exercise of competence-competence may, unfortunately, lead to arbitrators being biased against the party who objects to their jurisdiction.⁸²⁰ Arbitrators may, as a result, indirectly

⁸⁰⁸ Sulser (n 712) 127.

⁸⁰⁹ Landolt (n 733) 513.

⁸¹⁰ Sulser (n 712) 127.

⁸¹¹ Antonias Dimolitsa, ‘Separability and Kompetenz-Kompetenz’ in Albert Jan van den Berg (ed), *Improving the Efficiency of Arbitration Agreements and Awards: 40 Years of Application of the New York Convention* (Kluwer Law International 1999) 217.

⁸¹² Landolt (n 733) 515.

⁸¹³ Graves and Davydan (n 753) 157.

⁸¹⁴ Park, ‘The Arbitrability Dicta in *First Options v. Kaplan...*’ (n 741) 139.

⁸¹⁵ Landolt (n 733) 515.

⁸¹⁶ Sulser (n 712) 119.

⁸¹⁷ Arbitration Act 1996, s.30.

⁸¹⁸ UNCITRAL Model Law on International Commercial Arbitration 1985, art 16 (1).

⁸¹⁹ [1995] 514 US 938.

⁸²⁰ Reuben (n 718) 823.

advantage the compliant counterparty.⁸²¹ Generally speaking, arbitrators may not be directly interested in the dispute and therefore competence-competence would not violate the natural justice prohibition against a man being a judge in his own cause.⁸²² However, because arbitrators' fees are usually determined by their time and effort, there is a real risk that competence-competence may violate this principle.⁸²³ Even without reliable evidence of partiality, the risk remains because "justice should not only be done, but should manifestly and undoubtedly be seen to be done."⁸²⁴ Unquestionably, then, the risk of partiality casts a shadow over competence-competence. Moreover, there can of course never be a truly impartial arbitrator; competence-competence may only worsen this problem.⁸²⁵ Relief may be found in the rights reserved by English law for courts to review the arbitrators' decisions on their jurisdiction, but this undermines the efficiency of the arbitration process.⁸²⁶

Competence-competence also gives arbitrators the power to determine the validity, existence, or operability of the main contract, if disputed.⁸²⁷ These are legal issues for which the competence of arbitrators to address them is not always without doubt.⁸²⁸ Justice Hugo Black's dissent in *Prima*⁸²⁹ stemmed from this point since arbitrators are usually non-lawyers; where they are lawyers or are competent to apply the law, they are not legally mandated to do so.⁸³⁰ Competence-competence, therefore, gives arbitrators the possibility of getting things wrong by evaluating disputes differently

⁸²¹ *Regina v Bow Street Metropolitan Stipendiary Magistrate and Others* [2000] 1 AC 119 [133] (Lord Browne-Wilkinson).

⁸²² *ibid.*

⁸²³ Sandra Synkova, *Courts' Inquiry into Arbitral Jurisdiction at the Pre-Award Stage: A Comparative Analysis of the English, German and Swiss Legal Order* (Springer Science & Business Media 2013) 67.

⁸²⁴ *The King v Sussex Justices* [1924] 1 KB 256, 259 (Lord Hewart CJ).

⁸²⁵ Adrian Vermeule, 'Contra Nemo Iudex in Sua Causa: the Limits of Impartiality' [2012] 122 Yale LJ 384, 395.

⁸²⁶ *Barcelo III* (n 775) 1130.

⁸²⁷ *Sulser* (n 712) 125.

⁸²⁸ *Prima* (n 721) 407.

⁸²⁹ *ibid.*

⁸³⁰ *Reuben* (n 718) 839.

from competent state court judges.⁸³¹ Moreover, the lack of substantive judicial review avenues against arbitral awards can leave parties without redress.⁸³²

In the US, parties to a mandatory arbitration directed by courts normally retain the right to resort to trial if they are dissatisfied with the arbitrators' decisions.⁸³³ Also, the tribunal's ruling on its jurisdiction may later be overruled by court.⁸³⁴ In these cases, arbitration serves merely as an expensive precursor to litigation.⁸³⁵ The effectiveness of competence-competence may be further marred by the courts' unclear⁸³⁶ and divergent pronouncements on it.⁸³⁷ The US Supreme Court's *dicta* in *First Options*, for example, to the effect that courts should only set aside an arbitrator's decision in narrow circumstances, has been criticised as being confusing.⁸³⁸

Nonetheless, if upheld, competence-competence serves its intended purpose of enabling parties to resolve all their disputes in one forum.⁸³⁹ Disputants themselves do have a role to play in ensuring this happens, however: they should clearly specify in their contract what powers arbitrators have and appoint arbitrators who are independent, competent (both with respect to the subject matter and the parties), and available to resolve their dispute.⁸⁴⁰

(iii) *Choice of Laws*

Separability entitles the parties to choose a different law to apply to the arbitration clause from that governing the main contract.⁸⁴¹ This may be necessitated where the law governing the main contract does not favour arbitration.⁸⁴² Indeed, the UK Court of Appeal in

⁸³¹ Park, 'The Arbitrability Dicta in *First Options v. Kaplan...*' (n 741) 139.

⁸³² Reuben (n 718) 823.

⁸³³ Park, 'The Arbitrability Dicta in *First Options v. Kaplan...*' (n 741) 140, footnote 9.

⁸³⁴ Barcelo III (n 775) 1130.

⁸³⁵ Park, 'The Arbitrability Dicta in *First Options v. Kaplan...*' (n 741) 138.

⁸³⁶ *ibid* 139.

⁸³⁷ Sulser (n 712) 120.

⁸³⁸ Park, 'The Arbitrability Dicta in *First Options v. Kaplan...*' (n 741) 139.

⁸³⁹ Rau, 'Everything You Really Need to Know about "Separability"...' (n 739).

⁸⁴⁰ Bernardini (n 732) 49 – 51.

⁸⁴¹ Leboulanger (n 742) 7.

⁸⁴² Landolt (n 733) 516.

Sulamerica gave priority to the parties' express choice of law, followed by their implied choice, in the three-phase approach formulated for determining the law applicable to the arbitration clause.⁸⁴³ The court concluded that the system of law governing the substantive contract also applies to the arbitration clause if the parties do not expressly make this choice and there are no factors suggesting a different choice.⁸⁴⁴ Although Chong J in *BCY v BCZ* agreed with the approach in *Sulamerica*, he also noted the opposing view that the parties should ordinarily be taken to have impliedly chosen the law of the arbitration seat as the law to apply the arbitration agreement, rather than the law governing the substantive contract.⁸⁴⁵

In practice, parties rarely choose a different law for the arbitration clause.⁸⁴⁶ In fact, no specific attention is given to it by drafters of international contracts.⁸⁴⁷ Bernadini highlights a) the focus by drafters on ensuring the effectiveness and validity of the arbitration clause, b) the low likelihood that the law governing the arbitration clause will become contentious, and c) time and cost implications as possible reasons for non-consideration of a separate law for the arbitration clause.⁸⁴⁸ It is, however, equally possible that the parties are not aware of this choice or that their lawyers have limited or irrelevant knowledge of the available choices.⁸⁴⁹

The failure by the parties to specify the law applicable to the arbitration clause results in the tribunal making this decision, which may disadvantage one or both parties.⁸⁵⁰ Ultimately, this

⁸⁴³ *Sulamerica Cia Nacional de Seguros SA and others v Enesa Engenharia SA and others* [2013] 1WLR 102.

⁸⁴⁴ *Ibid* 114 [26].

⁸⁴⁵ [2016] SGHC 249 [48, 54].

⁸⁴⁶ Frédéric Bachand, 'Further Thoughts on Sulamerica: What About Transnational Rules?' (*Kluwer Arbitration Blog*, 20 September 2012) <<http://kluwerarbitrationblog.com/2012/09/20/further-thoughts-on-sulamerica-what-about-transnational-rules/>> accessed 06 April 2016.

⁸⁴⁷ Bernadini (n 732).

⁸⁴⁸ *ibid*.

⁸⁴⁹ Chatterjee (n 768) 539.

⁸⁵⁰ Jacqueline Chaplin, 'Hold on to your seats! A settled test for the proper law of arbitration clauses?' (*Kluwer Arbitration Blog*, 23 March 2012)

limits the freedom of choice that arbitration is grounded upon.⁸⁵¹ Yet, the law cannot be blamed if it provides rights which parties do not exercise.

(iii) *Termination of the Main Contract*

Separability means that termination of the main agreement does not automatically mean termination of the arbitration clause.⁸⁵² The tribunal would still have to determine the status of the main contract, which would be impossible if the arbitration clause fell with the main contract.⁸⁵³ If it is true that the parties do not give thought to the arbitration clause when forming the main contract, then it must also be true that they do not consider it separately at termination of the main contract.⁸⁵⁴ Again, a failure by the parties to exercise their rights cannot be blamed on the law.

IV. Application of the Separability Doctrine

(i) *Voidable and Void Contracts*

The distinction made by states between void and voidable contracts is one of the factors that has significantly influenced the application of separability.⁸⁵⁵ Due to this distinction, the US courts' approach to separability was, until recently, so narrow that it frustrated the true rationale for separability.⁸⁵⁶ Indeed, except in a few cases, separability was only applied to voidable contracts and not void contracts.⁸⁵⁷

A voidable contract is one which an aggrieved party is entitled to invalidate or set aside because the party was an infant when he signed it; or he would not have entered into the contract

<<http://kluwerarbitrationblog.com/2012/03/23/hold-on-to-your-seats-a-settled-test-for-the-proper-law-of-arbitration-clauses/>> accessed 08 April 2016. Links should not be underlined.

⁸⁵¹ John J. Broderick, 'The Future of Arbitration in the Settlement of International Commercial Disputes' [1984] 2 Notre Dame Int'l & Comp LJ 171, 174.

⁸⁵² Sulser (n 712) 122.

⁸⁵³ *ibid.*

⁸⁵⁴ Bernardini (n 732).

⁸⁵⁵ Redfern and Hunter (n 714) 106.

⁸⁵⁶ Monestier (n 729) 223.

⁸⁵⁷ *ibid.*

had it not been for fraud, mistake, or duress; or there is a breach of a material term or other justifiable ground to put the contract to an end.⁸⁵⁸ By contrast, a void contract is one which as a matter of law never officially exists,⁸⁵⁹ because it is invalid right from the outset.⁸⁶⁰ The rationale for applying separability only to voidable contracts was that there could be no arbitration clause separable from a contract that did not come into existence in the first place;⁸⁶¹ that nothing could come of nothing.⁸⁶²

In the US, this position was founded on the Supreme Court's decision in *Prima* in which the appellant alleged fraud by the respondent in making the contract.⁸⁶³ It was held that the fraud allegation affected the main contract generally and not the arbitration clause specifically.⁸⁶⁴ Therefore, there was no need for judicial intervention to verify the making or performance of the arbitration clause, and the arbitrator was justified to decide on the status of the main contract.⁸⁶⁵

The distinction between void and voidable contracts is arguably artificial: "in either case, the agreement of the parties is not legally binding."⁸⁶⁶ Likewise, in both cases, the test whether there was a proper agreement between the parties is applied to circumstances existing at the time of the making of the contract. Consequently, it is not always easy to distinguish void and voidable contracts.⁸⁶⁷ Furthermore, subsequent repudiation of a voidable contract by an innocent party means there is no contract to work with at all, as is the case with a void one.

⁸⁵⁸ *Three Valleys Municipal Water District v Ef Hutton & Company Inc J* [1991] 925 F.2d 1136.

⁸⁵⁹ Monestier (n 729) 233.

⁸⁶⁰ Jeffrey Thomas Ferriell, *Understanding Contracts* (3rd edn, LexisNexis 2014) 11.

⁸⁶¹ Monestier (n 729) 233.

⁸⁶² Schwebel (n 740) 2.

⁸⁶³ *Prima* (n 721).

⁸⁶⁴ *ibid.*

⁸⁶⁵ *Prima* (n 721) 402-404.

⁸⁶⁶ Leboulanger (n 742) 14.

⁸⁶⁷ Bergeron (n 731) 424.

In the light of these considerations, it is illogical for separability to apply only to voidable contracts. The US Supreme Court thus eliminated the distinction between void and voidable contracts in *Buckeye* to allow for the broad application of separability.⁸⁶⁸ UK courts were generally quicker in extending separability to void contracts than most American courts.⁸⁶⁹ In *Harbour*, for example, the UK Court of Appeal upheld the validity of the arbitration clause despite the main contract being void for illegality.⁸⁷⁰ This decision was later affirmed by both the UK Court of Appeal and Supreme Court in *Fiona*.⁸⁷¹ *Fiona* held that bribery in the formation of the main contract did not invalidate the arbitration clause merely because it was part of a tainted contract.⁸⁷²

A contract may be void on grounds of illegality imposed by statute or at common law.⁸⁷³ This raises public policy questions⁸⁷⁴ because illegality mainly concerns the character of the contract in which the arbitration clause is included, rather than the willingness of the parties to arbitrate.⁸⁷⁵ In *Beijing Jianlong Heavy Industry Group v Golden Ocean Group Ltd & Ors* ('*Beijing*'),⁸⁷⁶ the court applied separability and held that an arbitration clause may be invalidated on public policy grounds, but only if such grounds specifically touch upon the clause.⁸⁷⁷ The court echoed the 'powerful commercial factors' reasoning in *Harbour* and found nothing in the arbitration clause requiring the performance of an illegal activity.⁸⁷⁸ However, the court in *Beijing* did not define public policy. This was most likely because of the term's vagueness, which has rendered unsatisfactory all attempts at

⁸⁶⁸ *Buckeye* (n 720).

⁸⁶⁹ Monestier (n 729) 234.

⁸⁷⁰ *Harbour* (n 792).

⁸⁷¹ *Fiona* (n 715) 1725.

⁸⁷² *ibid.*

⁸⁷³ Turner, *Unlocking Contract Law* (n 766) 263.

⁸⁷⁴ *ibid* 278.

⁸⁷⁵ Chris Turner, *Contract Law* (3rd edn, Taylor and Francis 2013) 359.

⁸⁷⁶ [2013] 1 CLC 906.

⁸⁷⁷ *ibid* [26].

⁸⁷⁸ *ibid* [46].

clearly defining it.⁸⁷⁹ The consequence of this lack of clarity is inconsistency in the application of the separability doctrine.

Although the extension of separability to void contracts deserves commendation, the doctrine has not yet realised its full potential because *Buckeye* and *Fiona* have left some questions unanswered. Indeed, these cases introduced new challenges some of which are examined below.

(ii) *Contract Formation Disputes*

The application of separability continues to be limited where one party denies “that a contract ever existed, not in the sense that it would be void *ab initio*...but that there was never any sort of meeting of the minds...”⁸⁸⁰ For instance, the US Supreme Court in *Buckeye* unequivocally stated that contract validity questions must be distinguished from questions about whether the parties ever reached an agreement, raising the following concerns.⁸⁸¹

(a) *Old Wine in a New Bottle*

The distinction made in *Buckeye* between contract formation and contract validity issues has been criticised as having simply replaced the previous distinction between void and voidable contracts, with an even more confusing distinction between invalid and non-existent contracts.⁸⁸² The Supreme Court adopted an unnecessarily narrow approach in *Buckeye* when it found that *Prima* (and therefore separability) only covers contract validity issues (void and voidable contracts) and not contract existence issues (forgery, incapacity, and lack of authority).⁸⁸³ Whether this decision was consciously made or not, the court closed one gap and opened another. It may be said that the courts are back to the

⁸⁷⁹ Kojo Yelpaala, ‘Restraining the Unruly Horse: The Use of Public Policy in Arbitration, Interstate and International Conflict of Laws in California’ [1989] 2 Transnat’l Law 379, 380-381.

⁸⁸⁰ Nicholas Pengelley, ‘Separability Revisited: Arbitration Clauses and Bribery; *Fiona Trust & Holding Corp. v Privalov*’ [2007] 24(5) Journal of International Arbitration 445, 449.

⁸⁸¹ *Buckeye* (n 720) footnote 1.

⁸⁸² Rau, ‘Separability in the United States Supreme Court’ (n 724) 17.

⁸⁸³ *ibid.*

drawing board – but only in a different way – because the effect of a contract alleged to be void *ab initio* and one alleged never to have been formed is essentially the same – non-existence.⁸⁸⁴ The UK Supreme Court is commended for ending the ‘non-existence’ debate and removing the void-voidable contract restrictions, thereby promoting separability.⁸⁸⁵

(b) *Assent/Agency Issues: Lack of Clarity on Distinguishing Between Attacks Specific to the Arbitration Clause and those to the Main Contract*

As demonstrated in Section II, the law provides for impeachment of the arbitration clause on grounds which specifically affect the clause.⁸⁸⁶ However, there is a lack of practical guidance on how to distinguish assent to the arbitration clause from assent to the main contract.⁸⁸⁷

In *Fiona*, Lord Hoffman recognised that there may be cases when an attack on the main contract amounts to an attack on the arbitration clause.⁸⁸⁸ Examples include circumstances in which one of the parties claims his signature was forged or that the person who signed on his behalf did not have the authority to do so, with the result that the claimant did not agree to anything in the disputed document, including the arbitration clause.⁸⁸⁹ Beyond this, however, Lord Hoffman did not clarify what makes the attack one aimed directly at the arbitration clause.⁸⁹⁰ Further, the Court of Appeal in *Fiona Trust*⁸⁹¹ only provided ‘a trace’⁸⁹² of this distinction by giving an example of a direct impeachment of the

⁸⁸⁴ Pengelley (n 880) 449.

⁸⁸⁵ *Fiona* (n 715) 1725.

⁸⁸⁶ Arbitration Act 1996, s.7; Federal Arbitration Act 1925, s.2.

⁸⁸⁷ Rau, ‘Separability in the United States Supreme Court’ (n 724) 16.

⁸⁸⁸ *Fiona* (n 715) [17].

⁸⁸⁹ *ibid.*

⁸⁹⁰ *ibid.*

⁸⁹¹ [2007] Bus LR 686 [24] (CA).

⁸⁹² Grant (n 713) 874.

arbitration clause on account of fraud, based on Rix J's *dicta* in *Credit Suisse*.⁸⁹³

At present, it can be assumed that the distinction between assent to the arbitration clause and the contract generally would depend on the language the party resisting arbitration uses in their statements of claim. One possible solution is for each party to write their names and sign against the arbitration clause to evidence their agreement to it specifically. A signature template could be inserted right after the arbitration clause, similar to the one used for the main contract. This would make it easier to prove that the parties consented to arbitrate as per the clause.

It is suggested that after *Buckeye*, the question of whether courts rather than arbitrators should determine a challenge to the main contract likely depends on whether the challenge relates to the party's assent to the main contract.⁸⁹⁴ "If so, the defence – such as lack of assent or fraud in the execution – is one that courts can decide. If not, the defence – such as fraudulent inducement or illegality – is for the arbitrator."⁸⁹⁵ The use of 'likely' connotes the uncertainty that arbitration scholars, parties, and practitioners, are burdened with in the absence of unequivocal guidance from the courts.

Worse still, the arbitration clause is often a last minute or 'midnight' clause to which very little thought, if any, is given by the parties.⁸⁹⁶ Its consideration usually happens when the parties are already psychologically committed to the technical merits of the contract.⁸⁹⁷ In fact, it is common to find that one party does not even know about the arbitration clause.⁸⁹⁸ In such circumstances,

⁸⁹³ *Credit Suisse First Boston (Europe) Ltd v Seagate Trading Co Ltd* [1999] 1 Lloyd's Rep 784.

⁸⁹⁴ Christopher R. Drahozal, 'Buckeye Check Cashing and the Separability Doctrine' [2009] 1 Yearbook on Arbitration and Mediation 55, 90.

⁸⁹⁵ *ibid.*

⁸⁹⁶ Sapna Jhangiani, 'Conflicts of Law and International Commercial Arbitration – Can Conflict Be Avoided?' [2015] 2 (1) BCDR International Arbitration Review 99, 100.

⁸⁹⁷ Rau (n 739) 190.

⁸⁹⁸ Gross (n 737) 123.

demonstrating agreement to arbitrate may be problematic, giving rise to two possibilities:

- a. The fall-back position for a party seeking arbitration would be to rely on consent to the whole contract, which prevents independent consideration of consent to the arbitration clause; and
- b. The party resisting arbitration may raise the ‘midnight clause’ argument as a delaying tactic by alleging that he neither considered nor agreed to the arbitration clause.

It is argued that in these cases, it would be logical to attribute the existence of the arbitration clause to the existence of the main contract.⁸⁹⁹ This reasoning, however, makes it difficult for separability to prevail regardless of the status of the main contract.⁹⁰⁰

(c) *Incapacity*

“Incapacity has long been a ‘defence’ to the enforcement of a contract formed by a minor or a mentally incompetent person.”⁹⁰¹ It operates to make a contract voidable⁹⁰² and does not go to the core issue of contract ‘formation’;⁹⁰³ rather, it acknowledges the existence of a contract which is unenforceable against the party without capacity.⁹⁰⁴

On the contrary, the US Supreme Court in *Buckeye* categorised capacity with assent and agency, which it described as matters touching upon the initial ‘formation’ of a contract, to be resolved by courts and not arbitrators.⁹⁰⁵ Ware anticipates that this may be a step by courts towards eliminating the current distinction

⁸⁹⁹ Monestier (n 729) 241.

⁹⁰⁰ *ibid* 246.

⁹⁰¹ Ware (n 725) 118.

⁹⁰² Ferriell (n 860) 12.

⁹⁰³ Ware (n 725) 119.

⁹⁰⁴ George A. Bermann, ‘The “Gateway” Problem in International Commercial Arbitration’ [2012] 37 Yale J Intl L 1, 34.

⁹⁰⁵ *Buckeye* (n 720) footnote 1.

between contract formation and contract validity issues, which would enable separability to apply to broader categories of disputes.⁹⁰⁶ However, because it is difficult to distinguish between contract formation and contract validity issues,⁹⁰⁷ the reclassification may have simply been one of the Supreme Court's erroneous decisions.⁹⁰⁸ Either way, the consequence of this reclassification is a further restriction on the scope of separability.

True autonomy of the arbitration clause requires no distinction to be made between contract formation and contract validity.⁹⁰⁹ Otherwise, arbitration scholars such as Rau and Ware, who as of 2006, had already engaged in a decade of dialogue over whether a signature procured at gunpoint touches upon contract validity or contract formation, will continue debating such issues.⁹¹⁰

It is noteworthy that the UK courts, unlike their American counterparts, have rejected the notion that nothing can come from nothing.⁹¹¹ As considered before, the UK courts have progressed further by emphasising that an arbitration clause should only be rendered non-existent if it is specifically attacked.⁹¹² US courts could follow this, especially in the light of the fact that the UK Supreme Court relied on precedents from the US, Germany, and Australia in *Fiona*.⁹¹³

⁹⁰⁶ Ware (n 725) 119.

⁹⁰⁷ Howard M. Holtzmann, Donald Francis Donovan, Christopher K. Tahbaz and Catherine M. Amirfar, 'National Report for the United States of America (2013)' in Jan Paulsson and Lise Bosman (eds), *ICCA International Handbook on Commercial Arbitration*, (Kluwer Law International 1984, Supplement No. 76, October 2013) 1.

⁹⁰⁸ Gross (n 737) 114.

⁹⁰⁹ The US and the UK could take a leaf from France because the French approach, though outside the scope of this paper, has been applauded as reflecting the highest level of autonomy of the arbitration clause from the main contract – see Philippe Leboulanger (n 742) 3.

⁹¹⁰ Ware (n 725) 125.

⁹¹¹ Pengelley (n 880) 445.

⁹¹² *Harbour* (n 792) 723 - 724 (Longmore J).

⁹¹³ John Townsend, 'Foreign Law and Uniformity in English Arbitration: *Fiona Trust v. Privalov*' [2009] 14 Unif L Rev 555, 559.

(iii) *Courts' Piecemeal and Reluctant Approach to Separability Issues*

(a) *Piecemeal Approach*

The fragmented approach used by courts in handling separability is another cause for concern. In *Buckeye*, despite drawing the distinction between contract formation and contract validity disputes, the US Supreme Court only addressed contract validity issues.⁹¹⁴ Regarding contract formation issues, the court only stated that they arise where there is lack of assent, authority, or capacity to contract.⁹¹⁵ Moreover, the court categorically stated that they would not consider contract formation issues in the case before them.⁹¹⁶

Buckeye presented an opportunity for a comprehensive consideration of both contract validity and contract formation issues, yet the court did not take advantage of it. This evasive approach in *Buckeye* is indicative of the practical difficulties the courts perceive with total divorce of the arbitration clause from the main contract. It may have also been the court's way of buying more time to think about the practicalities of absolute separability. Also, some court opinions are not expressly stated and are left to the imagination of readers.⁹¹⁷ A classic example is *First Options*, in which it was held that courts, and not arbitrators, must look into allegations that no container agreement was ever formed due to assent or agency arguments.⁹¹⁸ Notwithstanding this, the court referred neither to *Prima* nor the word 'separability' in spite of their relevance to the decision.⁹¹⁹ Furthermore, *Buckeye* made no reference to *First Options*, even though matters pertaining to assent and agency, which *First Options* dealt with, were highlighted by

⁹¹⁴ *Buckeye* (n 720) footnote 1.

⁹¹⁵ *Ware* (n 725) 112.

⁹¹⁶ *Buckeye* (n 720) footnote 1.

⁹¹⁷ *Ware* (n 725) 112.

⁹¹⁸ *First Options* (n 819).

⁹¹⁹ *Ware* (n 725) 114.

the court in *Buckeye*.⁹²⁰ A key question is whether it is necessary for courts to reiterate or refer specifically to earlier decisions or principles. The need for clarity, consistency, and workability of separability yields an affirmative response to this question.⁹²¹ A piecemeal rather than an all-inclusive approach has therefore resulted in a confusing body of case law.⁹²²

Additionally, the court in *Prima* mainly relied on Section 4 of the US-FAA and then shifted its focus to Section 2 in *Buckeye*, yet both cases raised separability issues.⁹²³ The court has thus been criticised for violating the ‘Whole Act’ rule of statutory interpretation, which requires courts to consider a statutory provision in totality with the whole Act.⁹²⁴ Ross, however, maintains that the ‘Whole Act’ rule is not just wrong but is systematically wrong because it is unclear and lacks empirical evidence to justify its application.⁹²⁵ This inevitably means that Reuben’s call for the abolition of separability on the ground that the court violated the ‘Whole Act’ rule when deciding *Prima* is not endorsed.⁹²⁶ Scholarly arguments notwithstanding, the need for coherence means the court is deprived of any justification for its piecemeal approach in interpreting separability laws.

⁹²⁰ *ibid.*

⁹²¹ Bermann (n 904) 4.

⁹²² *ibid.*

⁹²³ Graves and Davydan (n 753).

⁹²⁴ Reuben (n 718) 842.

⁹²⁵ Stephen F. Ross, ‘Statutory Interpretation as a Parasitic Endeavor’ [2007] 44 San Diego L. Rev 1027, 1030.

⁹²⁶ Reuben (n 718) 842.

(b) *Possible Reluctance*

It took the US Supreme Court nearly 40 years after its first decision, *Prima*, to decide in *Buckeye* that separability applies both to void and voidable contracts.⁹²⁷ In the UK, *Harbour* and *Fiona* were approximately 15 years apart. In all four cases, the superior courts upheld separability by deciding that fraud,⁹²⁸ illegality,⁹²⁹ and bribery⁹³⁰ allegations against the main contract do not necessarily invalidate the arbitration clause.

Samuel argues that this time lag might be explained by the fact that the courts were not promptly faced with cases requiring them to address separability issues such as illegality.⁹³¹ This does not, however, explain the US Supreme Court's conscious decision not to consider contract formation issues in *Buckeye* when it clearly had the opportunity to do so.⁹³² Whatever the reason, the time lag between these seminal separability decisions only heightens concerns that the current legal uncertainties will continue into the future.

(iv) *Assignment of the Main Contract*

There are currently no specific statutory provisions within common law systems on what assignment of the main contract means for the arbitration clause.⁹³³ This gap has resulted in conflicting views by courts and scholars on the issue and reveals that separability has not yet been fully embraced.⁹³⁴ Consistent application of separability dictates that assignment of the main contract should not automatically result in assignment of the arbitration clause since they are distinct agreements.⁹³⁵

⁹²⁷ Leboulanger (n 742) 12.

⁹²⁸ Issue in *Prima* (n 721).

⁹²⁹ Issue in *Buckeye* (n 720) and *Harbour* (n 792).

⁹³⁰ Issue in *Fiona* (n 715).

⁹³¹ Samuel, 'Agora: Thoughts on Fiona Trust' (n 728) 496.

⁹³² *Buckeye* (n 720) footnote 1.

⁹³³ Leboulanger (n 742) 14.

⁹³⁴ *ibid.*

⁹³⁵ UNCITRAL Model Law on International Commercial Arbitration, art 16 (1); Federal Arbitration Act 1925, s.2; Arbitration Act 1996, s.7.

In *National Iranian Oil Company v Crescent Petroleum Company International Ltd, Crescent Gas Corporation Ltd* ('*National Iranian*')⁹³⁶ for example, the parties agreed that neither could assign their "contract or any rights and obligations [t]hereunder" without the prior written consent of the other party.⁹³⁷ The Commercial Court decided that the claimant had given the first defendant its consent to assign the contract to the second defendant, and therefore that the latter was a legitimate party to the arbitration.⁹³⁸

National Iranian raises two concerns. Primarily, the court treated the assignment of the main contract as automatically resulting in the assignment of the arbitration clause. In effect, the second defendant was bound to arbitrate. The court had, however, acknowledged that Section 7 of the UK-AA treats the arbitration clause as a separate agreement.⁹³⁹ The Model Law also provides that an arbitration clause in a contract is "an agreement independent of the other terms of the contract."⁹⁴⁰ This means that the arbitration clause cannot be treated as one of the ordinary rights or obligations which are transferred with the main contract upon assignment. Moreover, the reference by the parties to 'this contract' meant the main contract and not the arbitration clause. This interpretation is supported by the court's own analysis of the consent to the assignment which it found related to the entire contract.

Secondly, the court found no evidence that the parties agreed to disregard separability in their circumstances.⁹⁴¹ The court also acknowledged that without such an agreement, the arbitration clause must be treated as being separate from the main contract.⁹⁴² Therefore, by allowing the arbitration clause to pass to the assignee as an automatic consequence of assignment of the main contract, the court's decision contradicted its interpretation of Section 7 of the UK-AA. For separability to be clearly and consistently applied, the better approach, with respect, would have been to read Section 7 of

⁹³⁶ [2016] EWHC 510 (Comm).

⁹³⁷ *ibid* [2] (Mr Justice Burton).

⁹³⁸ *ibid* [30] (Mr Justice Burton).

⁹³⁹ *ibid* [9] (Mr Justice Burton).

⁹⁴⁰ UNCITRAL Model Law on International Commercial Arbitration, art 16 (1).

⁹⁴¹ *National Iranian* (n 936) [14].

⁹⁴² *ibid*.

the UK-AA as covering all matters concerning the fate of the arbitration clause, including its assignment. Better yet, specific consent to the assignment of the arbitration clause would prevent an assignee from resisting arbitration on grounds of lack of consent.⁹⁴³

Automatic assignment of the arbitration clause is argued to be pro-arbitration because it ensures that the initial right to arbitrate is not affected.⁹⁴⁴ This is contrary to the ordinary application of separability which only causes confusion.⁹⁴⁵ It may also be against the wishes of the parties as in the case where facts leading to the dispute occur before the assignment, implying that the assignee is not as well-placed to address the resulting issues as the assignor. A problem may equally arise where the contract is predominantly of a personal nature and requires specific performance by the original parties.⁹⁴⁶ Such instances justify specific consent to the assignment of the arbitration clause.

The arbitration clause is usually a short provision within a much larger document.⁹⁴⁷ Separate documentation of its assignment is therefore unlikely to require too much consideration, time, or effort, and would only minimally increase the cost of arbitration, if at all. Separate documentation of the assignment of the arbitration clause is thus an option worth considering. Legislatures could also reformulate Section 7 of the UK-AA and Section 2 of the US-FAA to include, in general terms, all matters pertaining to the status of the arbitration clause, like its assignment. At present, these sections only cover the validity, existence, effectiveness, irrevocability, and enforcement of the arbitration agreements.⁹⁴⁸ In the meantime, ‘assignment’ should, for example, be read into the terms ‘existence’ or ‘validity’; after all, a party may allege non-existence or invalidity of an arbitration clause because of the assignment.

⁹⁴³ Landolt (n 733) 519.

⁹⁴⁴ *ibid* 518.

⁹⁴⁵ *ibid* 518–519.

⁹⁴⁶ *ibid* 518.

⁹⁴⁷ Reuben (n 718) 838.

⁹⁴⁸ Arbitration Act 1996, s.7; Federal Arbitration Act 1925, s.2.

(v) The Big Picture: Whether Separability Should Be Abolished

These problems have equipped opponents with arguments against separability⁹⁴⁹ on grounds that it is “archaic, unworkable, and broader than necessary to accomplish its legitimate policy goals.”⁹⁵⁰ Heeding such calls without careful thought about the big picture would be an uncalculated move given the doctrine’s contribution to arbitration.

Today, several contracts include arbitration clauses.⁹⁵¹ This is because parties are confident that arbitration clauses are reliable and will survive the weaknesses of the main contract.⁹⁵² Such trends are a good indication that separability is serving its purpose: to uphold the will of the parties to arbitrate⁹⁵³ and not for legal scholars to raise debates, though such debates are crucial for developing the law. There has also been a notable increase in arbitration cases over the years. The London Court of International Arbitration (‘LCIA’), for instance, reported a ten percent increase in its cases in 2015 compared with 2014.⁹⁵⁴ Similarly, the American Arbitration Association’s (‘AAA’) arbitration claims increased from 101 in 1980 to 1,356 in 2006.⁹⁵⁵ The positive trend in the US has specifically been attributed to Section 2 of the FAA, which protects the arbitration clause against nearly any challenge by giving it a status independent of that of the main contract.⁹⁵⁶

⁹⁴⁹ Taylor Payne & Richard Bales, ‘What A Contract Has Joined Together Let No Court Cast Asunder: Abolishing Separability and Codifying the Scope of the Provisions of Arbitration Agreements’ [2017] 120 West Virginia Law Review <<https://ssrn.com/abstract=2931171>> accessed 28 April 2017.

⁹⁵⁰ Reuben (n 718) 827.

⁹⁵¹ Jan Paulsson, ‘Why Good Arbitration Cannot Compensate for Bad Courts – *Freshfields Hong Kong University Arbitration Lecture*’ [2013] 30(4) Journal of International Arbitration 345, 349.

⁹⁵² *ibid.*

⁹⁵³ Schwebel (n 740) 4.

⁹⁵⁴ London Court of International Arbitration Registrar’s Report 2015, 1. The 10% increase comprised 326 arbitrations and 6 requests for other forms of ADR. These 6 cases are insignificant hence it is appropriate to state that the 10% increase significantly arose from arbitration.

⁹⁵⁵ Born and Miles (n 726).

⁹⁵⁶ Gross (n 737) 113.

Furthermore, the choice of arbitration is an exercise of freedom of contract.⁹⁵⁷ When parties opt for arbitration instead of court, they are deemed to have consented both to the risks and benefits of arbitration.⁹⁵⁸ The risks include the challenges associated with the application and impact of separability, some of which have been evaluated in this paper. The benefits, on the other hand, comprise flexibility, informed choice, accessibility, cost and time efficiency, ease of enforcement of awards, and the neutrality and technical expertise of arbitrators.⁹⁵⁹ The upward trend in arbitration both in the US and the UK implies that parties perceive the benefits of arbitration to outweigh the challenges relating to the application and impact of separability.

In recent years, courts have also strongly advocated separability because it “produces more sensible results than not having it.”⁹⁶⁰ The UK Supreme Court spent over 398 years deliberating on the nature of arbitration clauses before it strongly decided in favour of separability in *Fiona*.⁹⁶¹ Without a doubt, none of the critics of separability have reflected on it for as much time as the courts. Thus, it would be hard to argue that their criticisms should take precedence over the courts’ views.

Besides, Ware acknowledges in his concluding remarks that his proposal for the repeal of separability has weaknesses.⁹⁶² Specifically, he cites the time and cost inefficiencies that would arise if separability were repealed.⁹⁶³ Acknowledging the doctrine’s benefits only weakens his arguments for its abolition. By all means, cost and convenience are important factors for businesses. This explains why the courts have relied on ‘powerful commercial reasons’ to uphold separability.⁹⁶⁴ The additional costs that parties would incur without separability would not only hamper domestic and international commerce, but also the international relations

⁹⁵⁷ Chatterjee (n 768).

⁹⁵⁸ Reuben (n 718) 823.

⁹⁵⁹ Broderick (n 851) 174.

⁹⁶⁰ Samuel, ‘Agora: Thoughts on Fiona Trust’ (n 728) 493.

⁹⁶¹ *ibid*.

⁹⁶² Ware (n 725) 132.

⁹⁶³ Stephen J. Ware, ‘The Centrist Case Against Current (Conservative) Arbitration Law’ [2016] 68 (5) Florida Law Review 1227, 1248.

⁹⁶⁴ *Harbour* (n 792) 725 (Hoffman LJ).

between states that arbitration has created.⁹⁶⁵ In other words, there are currently no compelling reasons, on balance, for separability to be repealed.

V. Conclusion

This paper attempted to assess the challenges relating to the application and impact of separability. It also sought to link separability to the popularity of arbitration and to provide a commentary on whether the challenges justify repeal of the doctrine. Returning to these questions, it is now possible to state that the benefits of separability outweigh the challenges associated with it.

It has been shown that there are powerful commercial reasons supporting separability, such as, *inter alia*, the need to honour parties' choice of arbitration instead of court and to provide them with a one-stop dispute resolution forum.⁹⁶⁶ Without separability, the public good associated with arbitration – easing state courts' workload – would thus be adversely affected.⁹⁶⁷ The will of the parties to contracts would also be undermined.

Furthermore, it has been established that separability is now broadly accepted, both by courts⁹⁶⁸ and legislatures.⁹⁶⁹ Key milestones in their acceptance include the abolition of the distinction between void and voidable contracts⁹⁷⁰ and, in the UK, the elimination of the need to take a literal interpretation of the words used by parties in the arbitration clause.⁹⁷¹ Despite these achievements, it is evident that there are still many threats to a

⁹⁶⁵ As of 17 April 2016, 156 states had subscribed to the Convention on the Recognition and Enforcement of Foreign Arbitral Awards (New York, 1958). The Convention is founded on reciprocity and hence facilitates friendly relations among member states. See: <http://www.uncitral.org/uncitral/en/uncitral_texts/arbitration/NYConvention_status.html> accessed 17 April 2016.

⁹⁶⁶ *Harbour* (n 792) 725 (Hoffman L.J.).

⁹⁶⁷ *Gross* (n 737) 117.

⁹⁶⁸ *Fiona* (n 715) [10].

⁹⁶⁹ Arbitration Act 1996, s.7; Federal Arbitration Act 1925, s.2.

⁹⁷⁰ *Buckeye* (n 720).

⁹⁷¹ Samuel, 'Agora: Thoughts on Fiona Trust' (n 728) 489.

consistent application of separability. These have mainly arisen from the lack of clarity and piecemeal approaches by courts on separability, leading to calls for its abolition.⁹⁷² Due to the limitations of the scope of this paper, challenges arising out of special categories of contracts (such as employment and consumer contracts) were not examined.⁹⁷³

Taken together, the evidence suggests that unless all stakeholders, including the courts, legislatures, and scholars, clearly and consistently endorse separability, its application and impact will continue to be restricted and will vary from case to case. Courts are particularly called upon to adopt an all-encompassing approach to addressing issues relating to separability, and to enhance the clarity and certainty of the doctrine. Rather than advocate the repeal of separability, scholars should formulate workable solutions to the current challenges.⁹⁷⁴ The fact that separability critics acknowledge loopholes in their own prepositions should motivate them to cooperate with advocates of the doctrine to realise this goal.⁹⁷⁵ Disputants must also play a role in upholding separability.

92 years have elapsed since the US-FAA was passed in 1925. It may now be appropriate for the legislature to consider amending it so as to expressly provide for separability and its crucial features, in accordance with Landolt's recommendations.⁹⁷⁶ This amendment could also save courts from the predicament of heavily relying on one section in one case and shifting focus to another section in another when the cases raise the same questions.⁹⁷⁷ It would also provide an opportunity to rethink the application of the ordinary rules of contractual interpretation to arbitration clauses.

Finally, considerably more research will need to be undertaken to determine, in quantifiable terms, the relationship between separability and the increase in arbitration cases.

⁹⁷² Ware (n 725) 121; Reuben (n 718) 838.

⁹⁷³ Reuben (n 718) 822.

⁹⁷⁴ Ware (n 725) 119.

⁹⁷⁵ *ibid* 132.

⁹⁷⁶ Landolt (n 733) 512.

⁹⁷⁷ Graves and Davydan (n 753) 158.

Separability has, to a certain extent, been clarified over the years, which has arguably led to the increased popularity of arbitration.⁹⁷⁸ The challenges with its application and impact have therefore been unnecessarily exaggerated by critics and do not, in all the circumstances, justify its abolition.

⁹⁷⁸ Gross (n 737) 113.

