Necessity and Coercion in Scots Criminal Law: A Critical Analysis

Submitted in fulfilment of the requirements of the degree of Ph.D.

Grant Barclay, LL.B. (Hons), LL.M.
School of Law
College of Social Sciences
University of Glasgow

December 2021

Abstract

This thesis provides a critical analysis of the substantive defences of necessity and coercion in Scots law. These are affirmative defences which do not deny the commission of the crime charged, but instead offer an explanation as to why the accused should nevertheless escape criminal liability. The approach taken to these defences in Scots law has tended to appear fragmented and unprincipled, with a strong focus on their limitations rather than their underlying rationale, such that the basis for exculpation is unclear. This in turn has ramifications for our understanding of the relationship between these two defences: necessity and coercion are seemingly distinguished by an arbitrary appeal to the nature of the threat and whether the accused 'chose' the crime they committed. With the above issues identified, this thesis proceeds with the following aims: 1) to conduct a historical (and comparative, where appropriate) review of necessity and coercion to discover what principles or factors shape the current rules for these defences in Scotland; 2) to explore the normative basis for exculpation when persons are forced to commit offences under extreme pressure and/or emergency, based on the emerging concepts of a constrained will and constrained choice; and 3) to determine how this normative understanding and division of the bases for exculpation might provide a more principled approach to the necessity and coercion defences in Scots law than presently exists.

This thesis concludes that necessity and coercion should exculpate based on a normative distinction focused on whether the actor suffered from a constrained choice (such that they tried to do the best in unfavourable circumstances) versus a constrained will (i.e. normal reasoning has been impaired by an emotional reaction, caused by the circumstances). An extensive analysis of emotions in criminal law defences is undertaken in support of this argument, highlighting that conduct undertaken in emotion does not vitiate capacity and is thus an appropriate site for moral (and hence legal) blame. This new framework for necessity and coercion is thereafter utilised to reconsider the core requirements of these defences in contemporary Scots law: that the threat be immediate; and that the threat be of death or serious injury. This re-examination reveals that the immediacy requirement in both defences should be replaced with one focused on the reasonableness of the response, and that the scope for coercion, understood as a constrained will, may be widened to include offences committed while suffering from a deprivation of liberty.

Acknowledgements

I would like to thank the following for their (direct and/or indirect) contributions to this project:

First and foremost, my parents Rona Sutherland and Bill Barclay. Without their constant support – continuing well into my adulthood! – I literally would not be here typing these words. My mother's belief in me started before I was born, and the strength of her conviction has had no small part to play in pushing me to ever loftier heights. I am grateful that they have always been supportive of my life choices, even if they like to mock me for being 'work shy'. With the conclusion of this PhD, those jokes will finally be put to rest (they will not be missed).

My wonderful supervisors, Professors James Chalmers and Fiona Leverick, who, in addition to the sage advice and expert knowledge one would expect from fantastic teachers of criminal law, provided many needed laughs and smiles along the way. I will never forget the classic line: "Maybe they should just take the hit?" (You probably had to be there). Extended research projects like PhDs can be horribly isolating, especially when conducted remotely, so I am eternally grateful that my monthly supervision meetings felt like bright spots among vast expanses of (social) darkness. Indeed, as I move forward with my budding academic career, I feel emboldened by the knowledge that I have their support and friendship. Thank you both.

The extended academic community that I interacted with during my research, who helped challenge my views and shape who I wanted to be as an academic myself. The creation of the Virtual Criminal Law Group during the pandemic provided me with a much-needed outlet for academic discussion and debate; I hope it continues to thrive even as I otherwise wish desperately for life to return to normal. To Rachel and JP: you both have been my rocks when I felt like I did not belong and my life, academic or otherwise, has been enriched by getting to know you. A particular mention should also be made to Professor Ben Berger at Osgoode Hall law School, York University, for providing mentorship and friendship when I needed it most. I arrived in Toronto with doubts about the value of my research – thanks to your kindness and passion I left with a fire in my belly that has never been extinguished.

I also cannot forget the generosity of the Scottish Graduate School for Arts and Humanities (SGSAH) in funding this project as part of their Doctoral Training Partnership programme. Not only did their financial support help enrich the quality of this thesis by allowing me to devote all my time to it, but their wonderful staff also helped ensure that my cohort and I had all the additional support we needed to develop as researchers. Through SGSAH funding I was able to collaborate on workshops held at the Royal Surgeon's Hall in Edinburgh, as well as a multiple day online conference to promote and celebrate interdisciplinary learning. Further, SGSAH funding allowed me to spend a month in Toronto, Canada doing research – a trip I will never forget. Special thanks go out to Professor Dee Heddon, Lindsay Wilson, and John Boyle for always being warm and welcoming during our interactions – in the case of Dee and John I also hope that life after SGSAH is treating you extremely well.

List of Abbreviations

The following abbreviations are used throughout the thesis:

Chalmers & Leverick, *Defences*: Chalmers, J & Leverick, F, *Criminal Defences and pleas in Bar of Trial* (Edinburgh: W Green & Son Ltd, 2006)

Gordon, *Criminal Law*: Gordon, GH, Sir, *The Criminal Law of Scotland,* Vol I (3rd ed, MGA Christie (ed), Edinburgh: W Green & Son Ltd, 2000)

Author's Declaration

I declare that, except where explicit reference is made to the contribution of others, that this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Signed: Grant Barclay

Contents

1. Introduction

In 1959, a doctoral student by the name of Gerald H Gordon wrote the following as part of his PhD thesis on criminal responsibility in Scots law:

> "There seems no reason for regarding legal rules as categorical… there is no reason… why in some cases other considerations should not allow a legal rule to be broken. The law is not the embodiment of absolute wisdom but merely a means of social control, and it would be socially disadvantageous, for example, to prevent the preservation of a building by action involving the theft of a ladder and a fire extinguisher."[1]

Sixty years and three published editions[2] later, the jurisprudence surrounding the substantive defences of necessity and coercion in Scots law has moved at a regrettably slow pace. Indeed, coercion and necessity were not formally recognised as substantive defences in Scots law until 1983[3] and 1997[4] respectively, and since then the reported case law has been sparse.[5] In 2001, Lord Justice-General Rodger suggested that the "result of so many years of neglect is that the contours of the defence of coercion or duress are not as sharply defined as those of many other aspects of our law".[6] Those same words could have been levelled at necessity. Twenty years on, the judiciary have, to some extent, been able to reflect on both coercion and necessity, formulating strict requirements about when a person might be fully exonerated for breaking the law.

However, one might also say that over the course of these limited opportunities the Scottish judiciary has slowly stripped down the normative foundation of these defences by building up impossibly strict rules of application based on public policy considerations: namely the rule of law and disincentivising civil disobedience. A healthy degree of

---

[1] GH Gordon, *Criminal Responsibility in Scots Law* (1960) PhD Thesis, at p.447 (discussing necessity). Accessed at: https://theses.gla.ac.uk/2753/1/1960gordonphd.pdf.

[2] Gordon's thesis would provide the basis for his seminal textbook, *The Criminal Law of Scotland*, currently entering its fourth edition. See L Farmer, "The Idea of Principle in Scots Criminal Law" in J Chalmers & F Leverick (eds), *Essays in Criminal Law in Honour of Sir Gerald Gordon* (Edinburgh University Press, 2010), pp.86-102, at p.98, fn.56.

[3] *Thomson v HM Advocate* 1983 JC 69.

[4] *Moss v Howdle* 1997 JC 123.

[5] In the 38-year period since *Thomson* in 1983 there have been fourteen reported appeals to the Scottish courts considering either a necessity (nine) or coercion (five) defence. Of that number, only two were successful (both necessity: *Tudhope v Grubb* 1983 SCCR 350; *Lord Advocate's Reference (No 1 of 2000)* 2001 JC 143) and in one the decision was overruled by a subsequent reference (*Lord Advocate's Reference*, *ibid*).

[6] *Cochrane v HM Advocate* 2001 SCCR 655 at 657.

stringency in relation to defences which can thwart or undermine legal prohibition is to be expected, and indeed should be praised, being necessary for the proper functioning of the criminal justice system. However, often the judicial discussion feels less like a balancing act, and more like a blunt imposition of one philosophy over another. It is less a discussion considering the competing principles constitutive of a liberal democracy, and more a tempering exercise intended to minimise the concept of permissible rule breaking as far as possible. Indeed, the reasoning of the courts has been limited to a reluctant acceptance of the underlying proposition that someone might intend to break the law and yet be blameless; a 'yes, but' never far from reach when discussing the existence of such defences.

This is a problem when, for example, the backbone of the legal protections for victims of human trafficking who are forced by their captors to commit crimes is based on the defence of coercion.[7] Human trafficking is not a new problem, and it does not appear to be going away any time soon, so it becomes imperative that our system of safeguards for such a delicate issue should be considered carefully. More generally, if necessity truly is the mother of invention, then it is important that the underlying framework is sufficiently robust to handle both the unmeritorious *and* meritorious claims as they arise. The criminal justice system fails when it takes the path of least resistance to inhibiting unmeritorious claims by enforcing arbitrarily strict rules and narrow grounds;[8] and currently it is failing.

The necessity and coercion defences are particularly hard to rationalise because they seem to defy traditional legal defence classifications. The academic literature on substantive criminal law defences which exculpate the actor for committing a *prima facie* crime generally splits such defences into two sub-categories: justifications and excuses. These concepts have been subject to nuanced interpretation over the years, but in their most basic form conventional wisdom holds that acts are justified, and actors are excused.[9] Specifically, a justification holds that the accused's act was, all things considered, permissible or even right, whereas excuse defences operate on the basis that while the act was wrong, the actor is nevertheless not blameworthy. Dsouza refers to this dichotomy between justifications and excuses as the 'wrongness' hypothesis.[10]

---

[7] Discussed in detail at section 9.5.
[8] Consider here Blackstone's ratio: "It is better that ten guilty persons escape than that one innocent suffer". (W Blackstone, *Commentaries on the Laws of England* (Vol. IV, Oxford: Clarendon Press, 1769), p.352.)
[9] For a helpful overview of these nuances, see M Dsouza, *Rationale-Based Defences in Criminal Law* (Hart Publishing: 2019), pp. 3-6.
[10] *Ibid*; see also at p.175: "a justification negates the objective wrongness of the prima facie offence, whereas an excuse negates the defendant's blameworthiness for causing the prima facie offence, whilst (usually)

If one were to try and formulate a rationale for necessity and coercion in Scots law, one might instead start with the hypothesis that what characterises both defences is a normative principle that sometimes it is understandable or even permissible to break the law. From this normative principle we might expect to generate legal rules. These rules would have to be considered and developed in the broader context of the criminal justice system, and they would require careful balancing against competing principles which reflect the multifaceted nature of criminal justice as a process in society.

Sadly, and as aforementioned, the Scottish judiciary have rejected this approach in favour of one which starts from the premise that the legal rules for necessity and coercion should be dictated by public policy considerations, lest the floodgates be opened and the defences become a kind of 'criminal's charter' to evade rightful punishment. Further, not only have the courts been reluctant to engage with the underlying normative values at play, they have often ignored them outright. The court in *Lord Advocate's Reference* suggested that necessity required no definition,[11] heavily implying that necessity was a situation where 'you know it when you see it' (and you only see it when these strict requirements are fulfilled).

This thesis therefore proposes to redress the balance. It is as much a plea for clarity as it is for reform, depending on how far one thinks these defences returning to their normative roots could be classed as a 'reform'. This thesis will argue that approaching questions of necessity and coercion in Scots law from their normative foundations will provide a more robust framework from which the criminal justice system can determine liability. The current framework provides insufficient answers for complex questions because the requirements created to limit the defences were not made with underlying normative values in mind. Questions about whether a victim of human trafficking can plead a defence of coercion, or whether necessity is an appropriate defence where an accused does not bring about a net positive result are difficult to answer under the current rules. This is because legal developments have emerged in an environment which is overly fearful of undermining legal integrity, thus undermining any broader notion of determining when society is comfortable saying that certain persons who commit offences should not be blamed.

---

leaving its wrongness intact". While space precludes an analysis of Dsouza's alternate, 'quality of reasoning' hypothesis, I believe that the theory put forward in Part II of this thesis is broadly compatible with both hypotheses.

[11] 2001 JC 143 at para [33].

With the above issues in mind, this thesis will explore a new, normative foundation for necessity and coercion by developing normative concepts which can act as helpful heuristics for determining the appropriate scope and requirements of defences like necessity and coercion, which I nominate as types of 'reactive defences'. I will focus on a division between necessity and coercion which is based on the concepts of trying to bring about a positive result in situations of emergency, on the one hand, and reacting to emotional and stressful stimuli on the other hand, respectively. Using this normative foundation, this thesis will examine several of the most controversial issues facing necessity and coercion in contemporary Scots law. There are limitations to this analysis. While the thesis aims to provide a broad outline for a normative approach to these reactive defences, it is not intended to be comprehensive. Space precludes analysing each requirement under the new framework, but the two most important aspects (temporal and threatened harm requirements) have been explored.

By my estimation, there is only one other issue that is controversial enough to also merit sustained discussion; whether necessity and/or coercion should be a defence to homicide offences. There are two primary reasons why this thesis does not examine the applicability of reactive defences to homicide offences. First, in a thesis which focuses on the normative aspects of these defences, it seemed that there was little to be said which would add anything of significance to the already voluminous literature on the topic. There are many legal and/or philosophical papers which examine killing out of necessity,[12] and some of the most compelling reform options, considering the current mandatory sentencing structure for murder, have already been made.[13] Second, and as aforementioned, space precludes a comprehensive analysis of every requirement under the normative understanding of necessity and coercion proposed here. To my mind, by exploring the temporal and threatened harm requirements in detail, this thesis leaves the normative theory expounded here in a more complete state, despite not claiming any kind of comprehensiveness. One can more readily 'plug in' the new understanding to the current

---

[12] See, for example, the series of papers about whether a net saving of lives should grant a permission to kill, stemming from Phillipa Foot's Trolley Problem hypothetical first posed in relation to the doctrine of double effect: P Foot, "The Problem of Abortion and the Doctrine of Double Effect" in *Virtues and Vices and Other Essays in Moral Philosophy* (Oxford: Basil Blackwell, 1978) (reprinted from (1967) 5 *The Oxford Review* 5) pp.19-31 at p.23; JM Taurek, "Should the Numbers Count?" (1977) 6(4) *Philosophy and Public Affairs* 293; JJ Thomson "Killing, Letting Die, and the Trolley Problem" (1976) 59(2) *The Monist* 204; "The Trolley Problem" (1985) 94(6) *The Yale Law Journal* 1395; "Turning the Trolley" (2008) 36(4) *Philosophy and Public Affairs* 359. (The Trolley Problem is discussed in more detail at section 5.2.3).

[13] Lord Kilbrandon's comments about removing the mandatory sentence or enabling coercion as a partial defence to murder are tried, tested, and still popular theories today. For whatever reason, legislative reform of the law of homicide and its accompanying defences has been a low priority for the successive Scottish governments: "Duress *Per Minas* as a Defence to Crime: I" (1982) 1(2) *Law and Philosophy* 185 at 193-4.

legal system, with the question of murder being left as more of a policy consideration than an essential aspect of the defences.

In addition to providing a normative understanding of necessity and coercion, this thesis provides a novel review of the current defences in Scots law. There is currently an absence of any wholesale analysis of necessity and coercion in the Scottish legal literature which, crucially, considers the defences in the context of their relationship with each other. During my research I have encountered no Scottish criminal law textbook (or from any other jurisdiction, for that matter) which attempts more than a cursory comparative exercise between the two defences. This is despite various examples from the case law where the two defences are treated as defences "without a relevant difference".[14] This thesis provides such a comparative analysis, including a more thorough justification for why such analysis is necessary. This thesis also includes several analyses of comparator jurisdictions – most notably English law in relation to the current requirements, and an examination of the legal understanding and response to emotions in England, South Africa and Canada.[15]

These comparative analyses are limited in several respects. First, the analysis of English law in relation to the current requirements is mostly limited to a discussion about the different forms of the necessity defence in English law, as well as mentions of English cases in relation to issues which have not arisen in Scots law. Generally speaking, English law has been used to supplement the analysis of the current Scots law requirements due to the lack of reported case law found in Scotland, it being a much smaller jurisdiction. English authority tends to be persuasive, though not binding, on Scottish courts. English cases can therefore provide some indication of the potential jurisprudential flow to be found in Scotland if an equivalent case did arise. Further, one core argument of this thesis is that the version of necessity received into Scots law is the English concept of duress of circumstances.[16] Perhaps here more than anywhere else, then, one might expect the Scottish courts to pay closer attention to the English approach in this area. In the limited cases we do have, there is little indication that the Scottish courts wish to deviate from the approach taken in English law.

Second, the comparative analysis of legal understandings of emotions in different jurisdictions is limited to substantive criminal law considerations. It does not therefore

---

[14] *R v Howe* [1987] AC 417 at 429, cited with approval in *Moss v Howdle* 1997 JC 123 at 127-8, which in turn was cited with approval in *Lord Advocate's Reference* 2001 JC 143 at 155, para [34].
[15] The preference for these jurisdictions is discussed below at section 6.2.
[16] See section 3.2.1.

consider how the respective criminal justice systems of these countries deal with the topic of emotions beyond positive legal rules. It does not, for example, consider how emotions might influence pre- or post-trial considerations, or jury deliberations.[17] It is limited to discussing the equivalent substantive defences which engage with emotional responses, for the purpose of determining how different jurisdictions take account of emotional reactions in the course of law breaking. All references to emotions in the law should therefore be understood within this context.

To summarise the above, the stated aims of this thesis are thus:

1) To critically analyse the current necessity and coercion defences in Scots criminal law;
2) With respect to the first aim, to present an account of necessity and coercion in Scots law whereby they can be regarded as variations of the same defence;
3) To formulate a normative understanding of reactions in law and determine how this understanding might influence the formation of reactive defences like necessity and coercion;
4) With respect to the third aim, to appropriately delineate between the kinds of reactions people experience in their lives – with attention paid to the dichotomy between people who act badly under intense emotions and those whose actions are better understood through positive intention and choice;
5) To provide a structured analysis of this normative understanding of reactive defences in the legal context.

With these stated aims in mind, Part I provides an analytical overview to help frame the rest of the thesis, utilising historical and comparative perspectives to provide greater context than the sparse modern case law provides. Chapter two provides a historical account of necessity and coercion in Scots law, examining the earliest forms of subjugation and compulsion understood at the time of Mackenzie in the late eighteenth century, through to the early nineteenth century and Hume's formulation of compulsion which would become the foundations for the modern coercion, and subsequently necessity, defences. Chapter three builds on this idea by explaining how necessity and coercion came

[17] On such an analysis, see e.g. S Karstedt, "Emotions and Criminal Justice" (2002) 6(3) *Theoretical Criminology* 299. Consider also the current European Commission funded project led by Stina Bergman Blix which seeks to contrast the judicial decision-making process in four different legal systems with varying 'emotional regimes' to determine differences and similarities in the construction of objectivity: *The Construction of Objectivity: An international perspective on the emotive-cognitive process of judicial decision-making (JUSTEMOTIONS)*, information available at: https://cordis.europa.eu/project/id/757625.

to be interlinked by Hume's treatment of compulsion in modern Scots law, as well as the Scottish Appeal Court's reliance on the English concept of 'duress of circumstances'. From there, the chapter begins an analytical review of the most important requirement for necessity and coercion in contemporary Scots law, that the accused must have acted from an immediate threat of death or serious injury. This review highlights several issues generated by the current characterisation/division of these defences – in particular issues of characterisation in relation to implied threats and issues of proportionality.

Chapter four concludes Part I by completing this analytical exposition, turning to focus more specifically on the temporal aspect of the immediate threat requirement, and its relationship with the secondary requirement that the accused was unable to resist the violence, also known as having no reasonable alternative to breaking the law, for a valid plea of either defence. Here attention is drawn to the historical approach in English law which placed a greater focus on the lack of alternative options as being determinative of a defence, providing for a comparative analysis with the current Scottish (and English[18]) approach(es) which views immediacy of harm as a necessary condition for exculpation. This strict adherence to an immediacy requirement imbues these defences with a strong degree of objectivity which might be regarded as counter-intuitive in many cases: the requirement seems to deny a defence when the rationale of these defences may otherwise point to there being one (one's will can be overcome despite a threat not being immediate). With this requirement firmly in mind, this chapter then goes on to discuss requirements of reasonable firmness, and what characteristics might be considered when determining the person of reasonable firmness for the purposes of identifying a qualifying threat. The last part of chapter four considers the remaining requirements which may have an influence on the credibility of an accused's plea, such as their prior fault and voluntary association with criminals, and provides a brief exploration of the current rationale of the defences, framed in the context of the justification/excuse distinction.

Part II focuses on exploring the normative foundation of reactive defences, focused specifically on the kind of conduct captured by the necessity and coercion defences, to try and formulate a broader normative proposition which might help frame the defences for legal exposition. Chapter five focuses on developing this initial, normative claim. What is it that the law seeks to exculpate when it admits a defence in these circumstances? Here the category of 'reactive defences' is developed, along with two normative concepts – situations of individual emergency and situations of extreme pressure – to help outline

[18] See now *R v Hasan* [2005] 2 AC 467.

precisely what kinds of reactions may warrant a substantive defence in law. Situations of individual emergency envisages those situations where a person attempts to create a net positive result whereas situations of extreme pressure cover those situations where persons, acting out of fear or other emotions, attempt to preserve themselves and their interests. I argue that these concepts can provide a pleasing division between necessity and coercion, based on the presence of a constrained choice or will respectively. The rest of chapter five explores theories of culpability in more detail, with particular attention paid to unpacking the idea of extreme pressure as constrained will and emotional responses as a site for normative blame, sufficient for legal blame and separate from the idea of choice. I explore different theories of culpability to determine why persons acting with a constrained will are not appropriate subjects for liability, apart from those who lack a reasonable choice.

Chapter six explores culpability for constrained will further by undertaking a comparative assessment of the legal system's understanding of emotions in substantive defences in Scotland, England, South Africa and Canada, with a subsidiary focus on these legal systems' approach to coercion/duress and necessity. This chapter seeks to examine how emotions have been implemented to discussions of culpability, and whether these discussions operate in the context of coercion/duress and necessity, or in other substantive defences (such as non-pathological incapacity for emotional stress in South Africa). Broad similarities can be drawn concerning the treatment of strong emotional reactions as incapacitating where recognised in law, such as instances of provocation or non-pathological incapacity, but tend to be mostly ignored in contexts (such as coercion or necessity) where what exculpates is based on the situation rather than the actor. The current legal landscape is therefore at odds with an understanding of reactive defences which accommodates exculpation for extreme emotional responses.

Chapter seven concludes Part II by moving on to evaluate theories of emotions in criminal law defences which are grounded in neuroscience. This evaluation is intended to bridge the gap between the purported constrained will which was identified in chapter five as an appropriate ground for exculpation, but seemingly rejected by the legal systems examined in chapter six. Different understandings of emotion are explored, including a mechanistic understanding which sees emotions as blind impulses which incapacitate, and cognitive understandings which see emotions as being based in affect and underlying moral values which are responsive to reason. Championing this second understanding of emotions, the chapter transitions to a discussion about how to incorporate such a normative understanding in a reasonably principled way for constrained will (such that it can provide

effective legal rules for a variety of actors), with an examination of reasonableness and how this might be construed. This thesis argues for both typicality of response and reasonableness of the reaction to ground a coercion defence, based on the context of the situation itself.

Finally, Part III of this thesis puts these concepts into practice by re-examining the two most important requirements of these defences from this new, normative perspective. Chapter eight examines temporal requirements in reactive defences from such a perspective, utilising the imminence requirement in justificatory self-defence and its rich literature as an analogy to examine their significance and meaning. I explore the contours of necessity as a broader principle of justice in reactive defences and its relationship with high probabilities of harm, a consideration which emerges as a separate, necessary condition for a defence to employing potentially lethal force, understood here as inevitability of harm. This inevitability of harm concept is shown to demonstrate the law's commitment to the sanctity of life. Within this perspective on temporal requirements as representing something other than pure necessity, we can draw conclusions about the relative importance of inevitability of harm in the context of situations of extreme pressure and individual emergency. The case is made that in necessity and coercion a greater focus should be placed on the options available to an accused, with the imminence of a threat providing credibility, but not being necessary.

Chapter nine re-examines the nature of the harm threatened, exploring the breadth of personal integrity as a concept capable of encompassing not just death or serious injury, but also other invasions of the person. Particular focus is placed on exploring deprivation of liberty and chronic pain as instances of non-trivial breaches of personal integrity, and their potential inclusion to the threatened harm requirement of reactive defences generally is explored. Utilising the normative conception of the defences put forward here, I conclude that deprivations of liberty may be sufficient to ground a plea of coercion understood as a constrained will: e.g. where victims of human trafficking are forced to commit crimes while under the control of their captors. This expansion of the threatened harm requirement, it is argued, is reasonable under the normative conception of reactive defences defended here, and tempered by the various other requirements still present in the law.

Indeed, despite all this thesis has to say about the basis of necessity and coercion, it is not intended to provide a complete framework. As such, this is not a law reform project. Rather, it is hoped that this thesis may promote a more thoughtful discussion about these

defences, and perhaps even provide the starting point for such a project. In that sense, this thesis might be seen as a rehabilitation exercise – a rebranding of sorts. Too long have these defences been either ignored or examined reluctantly, such that a deep analysis is sorely lacking. The ethos of this project is that sometimes persons have no option but to confront and reason through a hard choice – it is time for the law to do the same.

PART I

2. A Historical Account of Necessity and Coercion in Scots Law

2.1 Introduction

This chapter provides a historical background for the substantive defences of necessity and coercion in Scots law, covering the period from the mid-1700s all the way to the modern understanding of the defences which solidified in the mid twentieth century. We begin with a brief look at how culpability was generally understood by the late 1700s, with close attention paid to the legal writers of this time, including Mackenzie, Forbes and Bayne. A lack of reported case law during this period means that the works of these writers are particularly helpful for discovering the structure of the law. We see that culpability was a very nebulous concept, with each writer treating aspects of agency and culpability in a very disparate way. Compulsion as a mitigating factor was limited to a series of special relationships where the coercer could be said to have had a prior, legal dominium over the accused – e.g. parent and child or master and servant. Likewise, a publication of the ancient legal text *Regiam Majestatem* dating from 1609 suggests that necessity may have reduced culpability where a man stole meat to satisfy his hunger, but no broader notion of the concept exists.

The chapter then traces how compulsion evolved during the time of Hume, representing the dominant view at the turn of the nineteenth century, where a broader understanding had emerged which recognised coercion between unrelated parties. This expansion was, however, intended only as a careful response to the civil unrest caused by the Jacobite rebellion, and thus focused on the subjection and force caused by gangs and rioters. To this end, Hume thought that the plea would be a difficult one in times of peace in a well-regulated society. Nevertheless, Hume recognised that such compulsion could exist and thus outlined four strict requirements which he saw as essential to the validity of such a plea. Hume was even less enthusiastic about necessity, which he saw as irreconcilable with the precepts of civil order and thus rejected outright. Hume's statement on compulsion would become its last major substantive development in Scottish legal history, until the twentieth century when the defences began to re-emerge in cases in the 1970s.

It is for these reasons that I conclude that the necessity defence was historically understood as a form of compulsion in Scottish legal consciousness, at some point breaking away and becoming a separate entity, in part owing to the influence of English jurisprudence, which is explored in the remainder of this chapter. This argument sets the foundation for the following chapters, where I argue that the form of necessity received into modern Scots law is merely a variation of coercion, or compulsion, rather than a novel defence in its own right.

2.2 Culpability in the Eighteenth Century

Criminal law theory was underdeveloped before the late seventeenth century when Sir George Mackenzie provided the first real attempt at a discussion of principles of liability in *Matters Criminal*.[1] This appears to be an appropriate place to start a historical account of the coercion and necessity defences, and indeed culpability generally, since before this time the criminal law resembled a scattered collection of offences with very little structure, often being based on the collections of court decisions by prominent lawyers of the time.[2] In contrast, from the time of Mackenzie onwards we can begin to see attempts at systematic approaches to the criminal law, including greater categorisation of offences, as well as attempts to identify and specify rules of general application, including rules as to who could commit crime (capacity), and rules on parties to the commission of crime.[3]

Culpability in Mackenzie's time is understood in very basic and general terms. As a result, we unsurprisingly do not find anything that remotely resembles the fleshed out special defences of necessity and coercion as understood in Scots law today.[4] Despite the absence of specific and individualised defences, the writers of this period nevertheless did recognise that an actor might be compelled to commit a crime in a way which did not warrant the imposition of criminal liability. This was mostly recognised in terms of an actor's capacity, with a general presumption that such capacity was present. Thus, *dolus* or a guilty mind was generally required for criminal liability,[5] but exceptions and/or leniency

---

[1] Sir George Mackenzie, *The Laws and Customs of Scotland in Matters Criminal* (OF Robinson (ed), Stair Society, 2012).
[2] See, e.g., Balfour, J, *The Practicks of Sir James Balfour of Pittendreich* (1754) (PGB McNeill (ed), Vol.21-22, The Stair Society, 1962) generally.
[3] L Farmer, *Making the Modern Criminal Law: Criminalization and Civil Order* (London: OUP, 2016), p.85.
[4] Which is saying something, considering how underdeveloped the defences are by modern standards.
[5] Mackenzie, MC, Part I, Title 1, 4 (all citations to Part I unless stated otherwise); W Forbes, *The Institutes of the Law of Scotland* (1730) (Edinburgh Legal Education Trust, 2012), Part I, Book I, Chapter I, 3; A Bayne,

appear to have been granted for children,[6] sleepwalkers,[7] drunks,[8] those that were permanently "furious" (i.e. mad),[9] and we see that it was even questioned whether corporations or similar entities were capable of committing crimes.[10]

2.2.1 Necessity in the eighteenth century

The necessity principle was nebulous in the eighteenth century, and references to it are sparse. One of the few references originates from an older law found in Sir John Skene's *Regiam Majestatem* of 1609 called *burthynsack*, which held that a person would not be guilty of theft where they took a calf or sheep or as much meat as they could carry on their back.[11] It has been asserted that the rationale for this exception was an acceptance of the necessity of the situation,[12] but the rule as it is stated in *Regiam Majestatem* says nothing to this effect. Indeed, the absence of such a rationale is confirmed by the fact that when Mackenzie later wrote about this law, he argued that it should be restricted so that it only applies where the theft was committed "to satisfie his necessity" and there was no reasonable alternative.[13] Forbes refers to the necessity of the situation as though it were a settled matter, but he may well have been agreeing with Mackenzie on this point.[14] Bayne goes a step further and explains that it is not so much the *act* of stealing to be considered in such cases, but rather the *cause* and *motive* which, in these cases of necessity is much more powerful, arising out of the cravings of hunger and survival.[15] When Hume considered the law in his *Commentaries*, he notes that the true meaning of the passage is unclear since it appears to cover any theft of calf, ram or meat, "whether the thief be or be not necessitous".[16]

It is also unclear what effect the law had on the liability of the offender, since the first paragraph appeared to absolve the accused of any punishment, but the second

---

Institutions of the Criminal Law of Scotland. *For the use of students who attend the lectures of Alexander Bayne, J.P.* (1730) (Gale ECCO, 2011), p.9.

[6] Mackenzie, *ibid* para. 5; Forbes, *ibid* Chapter II, 2[1.].

[7] Mackenzie, i*bid* para. 6; Forbes, *ibid*.

[8] Mackenzie, i*bid* para. 7; cf. Forbes, *ibid* who is rather strict in rejecting any kind of defence or leniency for drunkenness.

[9] Mackenzie, i*bid* paras 8a & 8b; Forbes, *ibid*.

[10] *Ibid* para. 9. On capacity generally, see also Forbes, *Institutes,* Part I, Book I, Ch. II, 2-3; Bayne at pp.13-15.

[11] Sir J Skene, *Regiam Majestatem: The Auld Lawes and Constitutions of Scotland* (Edinburgh: Thomas Finlason, 1609), Book IV, Chapter 16, p.70.

[12] Chalmers & Leverick, *Defences*, para. 4.04.

[13] Mackenzie, MC, Title 19, 3.

[14] Forbes, *Institutes,* Part I, Book 4, Ch.X, 1.

[15] Bayne, *Institutions,* p.122-123.

[16] Hume, i, 54.

paragraph makes it clear that the thief should be "scourged".[17] *Burthynsack* may therefore represent one of the earliest examples of a necessity principle operating to reduce criminal culpability, as traditionally believed, or it may have been a more archaic concept based on some kind of *de minimis* exception which was adapted by later writers to suit the emergence of this concept in their time.[18] Either way, we can see that by the time of Mackenzie Scots law did recognise certain circumstances that might reduce or eliminate culpability, based on a lack of criminal intent. Nevertheless, constructive liability seems to have been prevalent, meaning that the relevant guilty mind was inferred more readily[19] in cases where there was an obvious wrongful act, akin to modern strict liability.[20]

2.2.2 The importance of special relationships to exoneration

Other than this nebulous concept of necessity which was, aside from the crime of theft, essentially non-existent, the concept of 'compulsion' can be regarded as the precursor to the modern necessity and coercion defences. This is due in large part to Hume's writing on the topic,[21] and its subsequent adoption into modern Scots law.[22] However, in the 1700s compulsion does not appear as a unique subject, with the word being used in laymen's terms to describe the concept of subjugation which finds equivalency with the modern law of superior orders.[23] As a result, almost all of the writing at this time which considers a person being forced to commit a crime does so from the context of art and part liability, with different forms of subjugation exculpating depending on the status of the subjugator, and/or the relationship between subjugator and subjugated.

---

[17] *Supra* fn.11. Hume appears to have been equally unclear about its impact on liability: see *ibid*. Cf. Hume's contemporary Alison, who states without doubt that burthynsack "did not amount to an absolute liberation from punishment of theft, but only a mitigation of its pains" and by the time of Hume and Alison had "long ago been abandoned" in favour of mitigation at the discretion of the judge: A Alison, *Principles of the Criminal Law of Scotland* (Edinburgh: William Blackwood, 1832), p.675.

[18] Cf PW Ferguson, "Necessity and Duress in Scots Law" [1986] Crim LR 103 at 104 who claims that there are no judicial authorities or institutional texts which support a necessity defence as even a mitigating factor only.

[19] Forbes, *Institutes*, Part I, Book I, Ch. I, 3: "But, because this is a secret Act of the Mind, Law infers it, in Some Cases from Conjectures, drawn from the Quality and Character of the Person, or the Nature of the Fact, or Circumstances attending it." See also Mackenzie, MC, Title 1, 4, para. 2.

[20] E.g. *Keay* (1837) 1 Swinton 543; MC, Title 1, 4: "some acts are so irregular of their own nature that the law requires only that the act be proved, without proving the dole or wicked design, as in sodomy, adultery, etc." Although cf. MC, Title 35, 3 where Mackenzie states that "in crimes we look to the design and not to the event". This seems to contradict his statement in a later paragraph of the same Title where he acknowledges that crimes might be committed without *dolus*: Title 35, 5 (and see below).

[21] Hume, i, 47ff.

[22] *Thomson v HM Advocate* 1983 JC 69.

[23] This appears to have equally been the case in England, see: PR Glazebrook, "The Necessity Plea in English Criminal Law" (1972) 30(1) *Cambridge Law Journal* 87 at 111.

Specifically, writings focussed on how far the command of the sovereign, a superior, husband or father might 'excuseth' the offender.[24]

Mackenzie stated that those assisting before or during the commission of the crime would be guilty of that crime art and part, and punished to the same degree as the principal.[25] This rule was qualified by the requirement that the assister knew that the assistance given was towards the commission of a crime; and this knowledge would not be presumed, but had to be proved.[26] Forbes also acknowledged that guilt art and part of a crime requires knowledge of that crime. Thus, "one accidentally present, who is merely passive, neither encouraging nor offering to hinder the Fact" would not be understood as art and part of a crime.[27] This would appear to extend to situations where one unknowingly assisted a criminal, such as where a person unwittingly helps a thief to escape with stolen goods.[28] Here the concept of remoteness, or in the later language of Hume "a backward and inferior part",[29] is introduced as Mackenzie points out that assistance must have had an 'immediate' influence on the crime since remote assistance, such as supplying weapons, was only punishable *paena extraordinaria* (i.e. arbitrarily, at the judge's discretion).

In his discussion of special relationships affording potential excusing conditions in Title 35, 5 of *Matters Criminal*, Mackenzie works down the social hierarchy of superiors and their subordinates, starting with a statement that the command of a Prince excuses altogether in lesser crimes, but in atrocious crimes it excuses only from the ordinary punishment (i.e. the death penalty), citing the proposition '*metus poenam attenuate non in totum tollit*' (fear lightens the penalty but does not lift it off totally).[30] This reference to fear might be regarded as a very early form of the 'concession to human frailty' logic that has come to permeate the modern law of excuse defences like coercion.

However, while Mackenzie's headline point is a relatively clear statement in itself, the accompanying verbiage here is confusing as Mackenzie suggests that mitigation of punishment was seen as appropriate in atrocious crimes because the accused in such a case commited the crime absent *dolo malo* and *qui citra dolum deliquit ordinaria poena non punitur & illi qui aliquid adversus suam voluntatem agit,* crimen non adscribitur *sed*

[24] Mackenzie, MC, Title 35, 5 & 6.
[25] MC, Title 35, 7, although cf. Title 35, 10 where Mackenzie suggests they are "punishable according to their proportional degrees of guilt".
[26] MC, Title 35, 7.
[27] Forbes, *Institutes*, Part I, Book I, Ch. I, 4.
[28] *Ibid*. Mackenzie gives the example of a person helping the principal to drive away another's cattle after the principal claims they are his own: MC, Title 35, 7.
[29] Hume, i, 53.
[30] MC, Title 35, 5.

cogenti (he who commits an offence without *dolus* is not punished with the ordinary penalty, and a *crime is not attributed* to the man who does some act against his own will but to the one who compels him).[31] It is hard to see why Mackenzie would point to the crime not being attributable to the accused as justification for why compulsion in cases of atrocious crimes would only mitigate the punishment – this reasoning would appear to point away from punishment completely, rather than at its reduction. Bayne agrees with Mackenzie's statement but does not repeat his verbiage,[32] and Forbes appears to be silent on the issue.[33]

The situation appears to have been similar for those compelled by magistrates at this time, although rather than state 'commands of a Prince or magistrate', Mackenzie curiously opts to treat each type of command separately, and for the latter states that: "[t]he command of the magistrate, acting as a magistrate or a public person, excuses or defends the committer from the ordinary punishment in atrocious crimes, and from all punishment in lesser crimes."[34] It is unclear why a distinction is made between such commands when, based on the text, the rules appear to amount to much the same thing.[35] Thus, commands from either Prince or magistrate would excuse from the ordinary punishment in atrocious crimes, essentially meaning that punishment for the crime would be decided arbitrarily by the sitting judge.[36] Likewise, commands from Prince or magistrate in lesser crimes would exonerate completely. In that sense, so far as compulsion could be considered a defence at that time, it was understood both as a complete and partial defence,[37] depending on the nature of the crime.

Mackenzie states that a mandate given by a master to his servant would excuse that servant from the ordinary punishment in atrocious crimes where "the master is known to be

---

[31] *Ibid*, my emphasis added.
[32] Bayne, *Institutions*, p12.
[33] Forbes, *Institutes*, Part I, Book I, Ch. II, 2[2.].
[34] MC, Title 35, 5.
[35] Bayne appears to confirm as much: *Institutions*, p.12.
[36] This was certainly preferable since the 'ordinary punishment' for atrocious crimes was generally capital, and we know that arbitrary punishment was "something less than death": OF Robinson, "Law, Morality and Sir George Mackenzie" in HL MacQueen (ed), *Miscellany Six* (Stair Society, 2009) pp.11-27 at p.13. On the peculiarities of ordinary and arbitrary punishment/crimes (in the context of assault), see G Barclay, "The Structure of Assault in Scots Law: A Historical and Comparative Perspective" (2017) University of Glasgow LLM(R) Thesis at 14-19 (accessible at http://theses.gla.ac.uk/8569/).
[37] An argument can be made to suggest that, rather than a partial defence, compulsion operated as a mitigating factor. I prefer the term partial defence as the circumstances of compulsion served to reduce the standard of punishment from ordinary to arbitrary – it therefore seems to find a closer analogy to the role of modern partial defences to murder which operate to reduce a charge of murder to culpable homicide which, functionally speaking, means a reduction of the penalty from a mandatory life sentence to one which is fixed by the judge, considering the circumstances of the case.

cruel".[38] There is no mention of lesser crimes, and thus one must assume that no such defence existed for servants obeying their masters, although an equivalent provision did exist for those under the orders of Prince or magistrate.[39] One would assume that where the law was only willing to admit excuses for certain types of crimes, excuses for lesser crimes would be the obvious candidate. Equally, in the final category which concerns familial relations, Mackenzie states that the command of a father or husband would excuse his wife or child in lesser crimes but not in atrocious crimes in and of itself.[40] It is therefore difficult to elucidate any kind of principle or policy which governed this area of law from his writings, and it is perhaps better thought that Mackenzie was here documenting the arbitrary rules as they developed through a smattering of cases.

2.2.3 Force as a necessary condition of compulsion

As aforementioned, the account of compulsion as a principle of law advanced by Mackenzie is one which more closely resembles the modern law of superior orders than any theory of coercion. Another aspect which demonstrates this point is that there is no indication that any force was necessary to convince the actor to commit the crime – rather the concession appears to be based on the notion that the actor felt compelled to act as they did because of the special relationship they maintained with the compeller, as either a child, spouse, servant or subject. Forbes, however, provides a contrasting analysis of situations where an otherwise innocent person might find themselves caught up in a criminal act. Forbes stands alone in suggesting that offenders are excusable when "compelled by mortal Threats, and irresistible Force, to commit a crime, or to be accessory to it".[41] This he calls "Subjection to the Power of others", which is clearly a recognition of circumstances more in line with the modern defences of coercion or necessity than with superior orders since the latter do not require there to be any risk of serious injury or harm to justify compliance.

In further contrast to his contemporaries, he also qualifies this excuse of 'subjection' by adding that the simple authority or command of a husband, father, or master alone would not justify a wife, son, or servant committing a crime in obedience.[42] Forbes therefore appears to be making the bold statement, in contrast to his

---

[38] MC, Title 35, 5.
[39] This position is confirmed by Bayne: *Institutions*, p.12.
[40] MC, Title 35, 6.
[41] Forbes, *Institutes*, Part I, Book I, Ch. II, 2[2.].
[42] *Ibid*.

contemporaries, that these special[43] relationships of subordination should not be considered as a mitigating factor in and of themselves, and that only a threat of danger to life or limb would be relevant to finding compulsion. This is certainly a more attractive account given how confused and contradictory Mackenzie's passages appear by comparison. However, that no general rule of exculpation existed for these special relationships without more is not inconsistent with the fact that, in a few recorded cases, there were circumstances where the relationship played an exculpatory role, albeit insufficient to exculpate in and of itself. Thus, although Mackenzie's apparent claim that special relationships could excuse in and of themselves can be doubted, perhaps such subordination was capable of excusing provided it featured something more than simple obedience.[44] Indeed, Mackenzie does suggest that the extent to which the mandate or command of a superior excuses was "variously debated by the doctors".[45]

If something more was required, then precisely what this amounted to was not made explicit. The prime candidate would be the 'mortal danger' espoused by Forbes, but the writing of Mackenzie does not suggest anything that drastic. A potential conclusion might be drawn from Mackenzie's inferential comment that a servant would be excused from committing an atrocious crime where the master was "known to be cruel",[46] and Hume even suggests that this should be understood as referring to "actual coercion" and a "reasonable fear of violence",[47] but even this can be questioned: 'cruel behaviour' is a far cry from the mortal danger that Forbes speaks of. Indeed, the example given is one where servants were treated leniently because their master was a 'robber' and they were unaware of this fact. Thus, the cruelty envisaged was understood as an objective characteristic of the master and had no relevance to the servant in terms of their culpability. In this sense, it is better thought of as an example of ignorance precluding liability.[48] We therefore have a strict requirement for mortal danger at one end of the spectrum, and an exclusion based on the prior character of the subjugator at the other.[49]

Nevertheless, Mackenzie is unequivocal in stating that in lesser crimes the commands of a father or husband could excuse and there does not appear to have been any

---

[43] Forbes is curiously silent on the topic of commands from a Prince or magistrate.
[44] Hume appears to have thought so: see Hume, I, 49: "where threats or violence have been employed by the husband to coerce her, a lower degree of terror shall excuse the submission of the wife".
[45] MC, Title 35, 5.
[46] *Ibid*.
[47] Hume, i, 49.
[48] See section 2.2.2 in relation to the concept of remoteness.
[49] Bayne might be taken to recognise the division of opinions here when he states, at *Institutions* p.12: "In leffer Crimes, the Command of the Prince or Magiftrate, *fome fay, of a Father*, excufe altogether" (my emphasis added).

requirement of force or danger as Forbes suggests. The conclusion is therefore that either Forbes or Mackenzie must be taken to be expressing an opinion rather than a description of the law at that time. A lack of organised case reporting makes it difficult to say with any great certainty which account was accurately representing the law, however on this issue Hume's *Commentaries* may provide some assistance. Mackenzie's proposition that the commands of a father mitigate is based on the authority of the case of *John Rae*[50] which was also reported in Hume, and makes no mention of force, only that the child complied with and aided his father in the theft of sheep, but was "assoilzied by the Court, in respect of his nonage".[51] Indeed, Hume acknowledges Mackenzie's view, suggesting that his use of this case as "proof of a more *general* position, That the command of the father entirely relieves the child from the guilt of the inferior offences" is unfounded, pointing out that *John Rae* was decided on the basis of the boy's age (or lack thereof), and that no other judgment exists which extends such a privilege to those under the command of a father.[52]

Mackenzie appears to base this general proposition from a second, Civilian source on which Hume is silent; a passage from the work of Jacobus Menochius in which he claims the issue is "fully treated".[53] Mackenzie cites the entire chapter, which is indeed based on higher commands generally, but Robinson correctly points out that only paragraph 41,[54] indeed a single sentence, is relevant to the command of fathers: "*Patris mandatum excusat filium* à *poena delicti leius,*[55] *non verò gravis* [The command of a father excuses a child for committing trivial crimes, but not for committing serious ones]."[56] The foundations for Mackenzie's view are therefore rather weak; a single line from a Civilian author accompanied by a singular Scottish case which was decided on a seemingly separate issue (or at least a very specific set of facts). It would therefore appear that subjection by a father in and of itself, if recognised under Scots law at all, was likely very rare and did not represent a settled principle in law. To this end, Forbes' requirement for mortal danger might be seen as preferable. If this is accurate, it may well represent one of the earliest elucidations of the 'immediate danger' requirement which has been received into modern Scots law on this topic.

---

[50] 1 January 1662, Hume, i, 48ff.
[51] Hume, i, 48-49.
[52] Hume, i, 49.
[53] MC, Title 35, 6, citing Menochius, *De arbitrariis judicum quaestionibus & causis*… (1613), *casus* 354.
[54] MC, Title 35, 6, fn. 23 in Sir George Mackenzie, *The Laws and Customs of Scotland in Matters Criminal* (OF Robinson (ed), Stair Society (2012)).
[55] The word '*leius*' is written in the 1630 edition of *De arbitrariis*… which seems most likely to be a typo for the word '*leuis*', which translates as of little consequence, unimportant or trivial: see PGW Glare, *Oxford Latin Dictionary* (London: Clarendon Press, 1968), p.1020, entry on "leuis", n.13.
[56] Menochius, *De arbitrariis judicum quaestionibus & causis*… (1607), *casus* 354, s.41 (my own translation).

2.2.4 A very special relationship: marriage

With regard to the specific situation of wives being under the compulsion of their husbands, by the 1800s Hume is seemingly aware of the argument that such circumstances could excuse in lesser crimes but rejects this view on the basis that it is more in line with the law of England, and also because he thinks it unwise to always assume that a wife who commits such an act in the presence of her husband does so against her will.[57] Curiously, Mackenzie relies on a paragraph from a medieval Scottish statute passed during the reign of William the Lion to hold that the defence would only operate in "lesser crimes",[58] but it is this same authority that Hume relies on to hold that both husband and wife are equally guilty, to be punished "according to their demerits".[59] An early adaption of the statute and paragraph(s) in question can be found in *Regiam Majestatum*. Paragraph eight states:

> "Albeit the wife fould obey hir husband, neuertheles in capitall or cruell crimes fcho is nocht oblifched to obey him. And fwa ilk ane of them fould be puniffed according to their demerites."[60]

This passage suggests that a woman was expected to obey her husband unless the commands would amount to a cruel or capital crime. In isolation, this appears to be in line with Mackenzie's view that compulsion would be recognised in lesser crimes – indeed one can comfortably argue that the second sentence concerning punishment is to be read in the context of the first, such that it is outlining the rules on punishment for cruel or capital crimes (in other words, only where the wife was liable art and part).

On this paragraph alone then, Hume's argument looks the weaker of the two. Hume also refers to paragraph seven of the statute, which states that:

> "gif the wife with the husband is convict, or confeffes that fcho was arte and part with him, they baeth fall be oblifched to anfwer… Bot quhen they are baeth participant in the crime, fwa they fall be partakers of the pane."[61]

What is particularly interesting is that Hume appears to omit the first sentence of paragraph seven when referencing the statute in his Commentaries. Indeed, the Latin variation of paragraphs seven and eight found in Hume roughly translates as:

---

[57] Hume, i, 48, citing W Blackstone, *Commentaries on the Laws of England*, Vol. IV (London: Clarendon Press, 1769), p.28.
[58] MC, Title 35, 6; William I, c.19.8.
[59] Hume, I, 47. Hume cites both paragraphs seven and eight of the statute (discussed below).
[60] Skene, *Regiam Majestatem*, Statutes of King William, p.6, chapter 19, para. 8.
[61] *Ibid*. para. 7.

> "However, since both were partners in crime, so too will they share the penalty. Though a wife should obey her husband, she should not obey in atrocious crimes. Thus both should be punished according to their own desserts."[62]

There is no mention of the prior condition that the wife should have been convicted or have confessed, found in the version provided by Skene. Similarly, in Balfour's *Practicks* when debating the subject he states:

> "*Gif the wife be convict or confes*, that in committing of ony crime or trefpas be hir husband fcho gave red, counfall or help, he and fcho baith may be accufit and punift thairfoir… feing thay be baith togidder guiltie, or baith togidder innocent, or zit the ane guiltie, and the uther innocent, the wife fould not be maid quyte and fre in all fic caufis".[63]

That Hume was aware of the above passage is made explicit by the fact that he refers to this passage in his *Commentaries*, but again he omits the prior text and begins by saying "the wyfe sould not be maid quyte and fri in all sic causis".[64]

This is a significant omission since it provides an important qualification to the liability of the wife by suggesting that it is only on admission or conviction that the wife can be said to share the punishment with her husband. From reading the primary sources there was clearly a degree of scepticism surrounding the claim that a wife was always to be treated as acting under the compulsion of her husband, but it is clear that there may have been situations where women were being exculpated on such a basis, and there does not appear to be any focus on a requirement for her to have been acting on a fear of 'mortal danger' as suggested by Forbes. Hume seems to concede as much in his analysis of the law when he states that the "utmost lenity, therefore, to which our Judges might incline, would be to excuse the wife for venial trespasses or petty crimes, to which her obedience of his orders may have constrained her".[65]

---

[62] Hume, i, 47.
[63] The Stair Society (ed), *Practicks of Sir James Balfour of Pittendreich Vol 1* (Vol 21, 1962), p.96 (my emphasis added).
[64] Hume, i, 47, fn.5.
[65] Hume, I, 49.

2.2.5 Summary

The period from the mid seventeenth to mid eighteenth century can therefore be regarded as rather unsettled, with principles of culpability either non-existent, or just beginning to form. What seems clear is that while certain circumstances of extreme pressure might arise that could provide grounds for an accused's acquittal, or mitigation of their sentence, this was generally on the basis of the special relationship an accused shared with a co-accused in such cases, and these tended to be decided on the basis of a subjection which shared a closer analogy to the modern law of superior orders. This is especially so in examples where no requirement of a threat of death or serious injury was necessary to sustain the defence. There, the existence of the special relationship itself can be seen to provide the rationale for mitigation of punishment, or even exoneration. Above all, *dolus* appears to have been essential for guilt, and this appears to have been more easily displaced if the accused could prove subjection to another. Put differently, if the accused could prove that their will had been overborne by their superior, they would have grounds for mitigation.

Necessity, if recognised at all, was limited to the niche topic of theft to satiate hunger. It may well be that other circumstances of necessity which did not involve theft or hunger provided mitigation, but a lack of documented reports at this time leaves this question unanswered. In any case, there appears to have been a reluctance to recognise a defence which would leave the criminal justice system without an appropriate target for punishment (to be contrasted with compulsion in special relationships where the subjugator could be targeted). We now move on to consider how these aspects of culpability developed into the nineteenth century by the time of Hume, one of the most influential writers in Scots criminal law, whose work still provides the basis for the modern defences of coercion and necessity.[66]

2.3 Culpability and Compulsion in the Nineteenth Century

As we saw in the last section, writings on the kind of compulsion which has come to be understood through the defences of necessity and coercion in Scots law seems to have been almost non-existent. This appears to have persisted until the time of Hume, who recognised a distinction between the subjection found in certain relationships, and the

[66] See, *infra*, chapters three and four generally, and specifically at section 3.2.1.

compulsion that could be brought to bear on an accused by strangers. This recognition can perhaps be explained by reference to the political events which took place between the publications of Mackenzie's *Matters Criminal* and Hume's *Commentaries*. Many of the cases that Hume cites took place around the time of the Jacobite rebellion, when Scotland was in a state of civil war.[67] Beyond the rebellion itself, the presence of rebel gangs at this time was not uncommon. There was thus an influx of cases where persons helping rebels claimed they were forced to comply for fear of the consequences.[68] It is perhaps for this reason that Hume felt it necessary to preface his discussion of compulsion between unrelated persons by saying that it was "reasonable to distinguish between situations of great commotion, or extensive danger, and the ordinary condition of a quiet and well regulated society".[69] He may well have feared that the plethora of favourable cases that had emerged during this time might be taken as evidence of a broader principle of exculpation which was, in his view, too lenient.

2.3.1 Compulsion in times of great commotion

Thus, in times of war or rebellion, when the "protection of the Government is suspended", allowances would be made to individuals who had to resort to "the great law of self-preservation to govern… their conduct for the time".[70] It seems that this rule operated as a form of pardon for those who had, in some way or another, aided rebels or gangs and thus to prevent charges of treason once the government of the day had re-established its authority and captured the offenders. It was likely thought disproportionate to charge persons with a capital crime for providing food, shelter or other forms of aid to rebels when they had no real alternative.[71]

Some comments can be made about the grounds for exculpation. Perhaps unsurprisingly, the name given to the defence at this time seems to have been unsettled, such that in the case of William Gilchrist the defence of "constraint and compulsion" was found relevant,[72] whereas James Graham was acquitted from aiding Rob Roy and his gang

---

[67] This is not to suggest that the political climate in the 17th century was any more settled, but rather to attempt to explain specifically Hume's reasoning.
[68] E.g. the case of *William and John Riddell*, indicted for treason after furnishing rebels with provisions and transporting their cannon in 1681: Hume, i, 50. See also the cases of *Andrew Fairney*, *James Purdie and others* (1720); *Roger Hews* (1720); and *Robert Main* (1725): i, 51.
[69] Hume, i, 50.
[70] *Ibid*.
[71] Hume provides the example of *William and John Riddell*, who were tried and acquitted on November 7, 1681 on charges of treason for providing rebels with provisions and help transporting their cannons: i, 50.
[72] *Ibid* at 51.

on the basis that "he was not possessed of arms to make resistance at the time" and thus the defence of "constraint by force" was fully made out.[73] This latter case suggests that there was a requirement for an inability to resist the violence. Chalmers and Leverick note that Hume makes no reference to a requirement for an explicit threat of violence unless the accused complies, suggesting that an implied threat would suffice.[74] We can probably assume the correctness of this position on the basis that Hume states, in his introductory statement to this type of compulsion, that some allowances should be made for a person who "yields [to rebels] against his will… in the reasonable fear of military execution if he refuse".[75] It is perhaps more accurate to state that, given the background circumstances, there was a broad and implied assumption that anything done in aid of rebels was done under compulsion, provided the accused could point to facts which supported this narrative.

Indeed, Hume stated that an accused may even be excused when they had spent a long time with the gang, provided they could prove an "outrageous compulsion", and that they escaped the situation at the first opportunity.[76] Thus, while no strict temporal requirement appears to have been recognised, it is clear that a criminal intention could be inferred from the duration of time spent with the gang, and whether the accused had capitalised on opportunities to flee.[77] Relatedly, Hume also stated that actions done by an accused had to have been "fairly imputable to the constraint of his situation", suggesting a causation requirement. Finally, Hume suggested that the defence might have its strongest claim in the case of those forced into piracy, since the compulsion would be "irresistible in such circumstances".[78] The implication appears to have been that, at sea, a person accosted by pirates would have nowhere to escape and thus no alternatives to joining them.

2.3.2 Compulsion in a well-regulated society

In contrast, Hume was sceptical about the availability of compulsion in times of peace, pointing out that where "every man is under the shield of the law, and has the means of resorting to that protection, this is at least somewhat a difficult plea".[79] Nevertheless, Hume did recognise that such a plea existed in the context of a well-regulated society, and

---

[73] *Ibid* at 52.
[74] Chalmers & Leverick, *Defences*, at para. 5.08.
[75] Hume, i, 50.
[76] *Ibid*.
[77] On this point, see also the case of *Robert Main*, October 11 1725, discussed in Hume at 51.
[78] *Ibid*.
[79] *Ibid* at 52.

may even serve to exculpate in 'atrocious' (i.e. serious) crimes, provided the following qualifications were met:

> "[A]n immediate danger of death or great bodily harm; an inability to resist the violence; a backward and an inferior part in the perpetration; and a disclosure of the fact, as well as restitution of the profit, on the first safe and convenient occasion."[80]

What is perhaps most curious about these requirements is their striking resemblance to the rough stipulations outlined by Hume in relation to compulsion in times of unrest. Indeed, on the basis of the requirements in times of unrest as Hume describes them there is reason to believe that the rules were in fact the same and that it was the standard to be met which differed between the two situations. The accused in both scenarios had to be acting to avoid an immediate threat to life or limb,[81] with an inability to resist or otherwise escape the violence, and must have taken the earliest opportunity to flee in order to plead the defence. Perhaps the clearest distinction that can be made between the two situations is the fact that Hume makes it clear that in times of disruption a person may nevertheless have the plea even where they spent some time with the rebels before being able to safely flee. This likely represents the reality that in the absence of a well-regulated society to rely on (i.e. an effective authority), what constituted a reasonable opportunity to escape would be very different. The distinction Hume draws between times of unrest and peace is one which has, despite its lack of clarity, been adopted into modern Scots law.[82]

2.3.3 Necessity in the nineteenth century

Broadly speaking, Hume's discussion of compulsion and the four requirements therein provide the foundations for the modern Scots criminal law defences of both coercion and necessity.[83] This has come to be despite Hume being very critical of a general necessity defence. As we saw above, *burthynsack* provided a very limited form of defence in situations where a person was starving and stole food to eat. Hume argued that a defence of 'compulsion by want' risked "subvert[ing] all security of property, by confounding the common notions of honesty among our people, and throwing into every man's own hand the estimation of his own wants and distresses".[84] At a theoretical level, it is interesting to

---

[80] *Ibid*.
[81] Alison states that the threats to property could suffice, but there is no recorded authority for this view and the passage containing this claim has not been adopted into the Scottish case law: Alison, p.672.
[82] *Thomson v HM Advocate* 1983 JC 69 at 78.
[83] *Thomson v HM Advocate* 1983 JC 69; *Moss v Howdle* 1997 JC 123. See section 3.2.1 below.
[84] Hume, i, 54.

note that Hume regarded necessity as we would understand it as a form of compulsion. However, and at a practical level, Hume thought it inappropriate and ultimately impossible to determine what counted as true necessity or whether the accused was genuine in their needs, as this would necessarily involve investigating "the whole history of the offender's life".[85] In this sense Hume appears to be touching on what has come to be known as the 'democracy problem',[86] rejecting the idea that courts should make any such value judgements in their capacity as impartial arbiters of justice. For Hume, such judgements were to be left to the other branches of government – namely executive clemency in the form of a Royal pardon.[87]

Alison appears to agree with Hume that "extreme distress or hunger" offered no legal defence, but stated that, where such facts were proved to exist, they could offer mitigation of punishment.[88] Alison relies on the old law of *burthynsack* for this proposition, pointing out that while the provision never fully exculpated, it nevertheless provided mitigation of punishment for such thefts. This law, he admits, "has long ago been abandoned", but suggested that mitigation in cases of severe want was left to the discretion of the judge in lesser crimes, with capital crimes subject only to royal mercy, as per Hume.[89]

2.3.4 Summary

The above discussion, and in particular the dichotomy between rules for compulsion in times of peace versus times of rebellion, reveals a tension between the primacy of the rule of law, on the one hand, and other principles of justice such as fairness, on the other. It is clear from Hume's writings that he was reluctant to accept a general necessity principle into the criminal law unless it was understood on the strictest terms, for

---

[85] *Ibid*.
[86] For a detailed account, e.g. S Robinson, "Necessity's Newest Inventions" (1991) 11(1) *Oxford Journal of Legal Studies* 125 at 132ff. See also F Stark, "Necessity and *Nicklinson*" [2013] *Criminal Law Review* 949 at 961-64.
[87] Hume, i, 55. Hume relies here on comments made by Blackstone about the beneficial nature of a monarchical form of government in being able to provide the necessary rigidity of legal rules, whilst also offering a more humane touch where required in unique circumstances in the form of official pardon. Blackstone himself appears to have also been ambivalent about a necessity defence, seemingly agreeing with Bacon's views that it could be pled in situations of hunger or where drowning persons were fighting for a plank, while also agreeing with Hale that theft under such circumstances was still a felony: see Glazebrook, *op cit*. at 110-11; Blackstone, *op cit* at pp.15 (stating that theft in the case of hunger is "far more worthy of compassion") and 186 (neutrally citing Bacon's plank scenario); A Norrie, *Crime, Reason and History: A Critical Introduction to Criminal law* (3rd edn, London: Cambridge University Press, 2014), at p.202.
[88] Alison, p.674.
[89] *Ibid*.

fear that it would undermine the law and its authority. Hume therefore rejected a defence for the starving thief, and it is unclear on which circumstances such a defence might be pled.

As aforementioned in the introduction to this section, many of the cases which Hume cites would have been tried during or after the rebellion, leading to a disproportionate influx of cases all suggesting that where an accused was placed under the compulsion of another, irrespective of any prior special relationship, they might have a complete defence. His writings on compulsion during peace time may therefore be seen as his attempt to restore balance and the primacy of the law to the benefit of civil order. His views on necessity and the law of *burthynsack* for example are particularly scathing, owing to the potential of such rules to throw the law into a subjective chaos where the individual can decide what is, or is not, necessary according to their own wants. Hume's views on exoneration by compulsion, cognisant of its potential destructive power, have stood the test of time and become the foundational framework for the modern coercion, and arguably necessity, defence(s).[90]

2.4 Compulsion in the Twentieth Century: From Compulsion to Coercion

Throughout the nineteenth century writers appear to have done little more than repeat Hume's account of compulsion when engaging with this topic. The notable Scottish writers on criminal law of this period, namely JHA Macdonald and later AM Anderson, cite his *Commentaries* heavily in their accounts. Indeed, the first four editions of Macdonald published between 1867 and 1929 rely exclusively on Hume and the cases he cites for authority.[91] Further, the fifth edition (1948) includes only one additional reference, to the English necessity case of *Dudley and Stephens*[92] for the proposition that compulsion by the pressure of extreme want does not provide a defence. Likewise, in Anderson's later account of compulsion in 1892 there are only references to Hume's *Commentaries* and *Dudley and Stephens*, and this is true of both editions.[93] If there were any unreported Scottish cases on compulsion decided during this time period, they were seemingly not notable enough to warrant a mention.

---

[90] See, *infra*, section 3.2.1.
[91] Macdonald, *A Practical Treatise on the Criminal Law of Scotland*, 1st edn (W Paterson, 1867), p.17; 2nd edn (W Paterson, 1877), p.13; 3rd edn (W Green & Son Ltd, 1894), p.13; 4th edn (W Green & Son Ltd, 1929), pp.14-15.
[92] (1884) 14 QBD 273, discussed in more detail below.
[93] AM Anderson, *The Criminal Law of Scotland* (Edinburgh: Bell & Bradfute, 1892), p.6-7; (1904), p.16.

2.4.1 Macdonald and Anderson on compulsion and Hume

Substantively, Macdonald's *Practical Treatise* devotes a singular paragraph to compulsion which is essentially an abbreviated restatement of Hume. Macdonald states that compulsion applies where a large mob force an individual to act "by absolute compulsion, or by the constraint of threats of death or serious injury",[94] and highlights that in addition to offences committed during large commotions, or of piracy on the high seas where no escape was possible would be excused, so too could there be cases where compulsion by an individual would suffice to ground the plea.[95] Indeed, very little appears to have changed in Macdonald's treatment of the topic from the publication of the first edition in 1867 to the fifth in 1948.[96] The lack of development demonstrated in Macdonald's account can, however, be contrasted with Anderson's treatment of the topic which demonstrates that broader conceptions of compulsion as a defence had emerged by the time of his writing in 1892. In addition to the usual threat requirement, Anderson adds the qualification that such threats must have been "of such a nature as to overcome the resolution of an ordinarily constituted person of the same age and sex as the accused".[97] It is curious that Macdonald should fail to mention this requirement in his own work when preparing the third edition two years later. Anderson cites no authority for this proposition, so there remains the possibility that he was merely expressing his own view or drawing on English authority, rather than stating the settled law of Scotland. In either case the requirement, as espoused by Anderson, has found its way into the modern law of coercion.[98]

Discussions of necessity during this period are even scarcer and, where they do appear, necessity is seen as derivative of something else. There is certainly no attempt to develop a broader conception of necessity as a special defence in the same vein as compulsion. Indeed, from an examination of the nineteenth century accounts we can conclude that the law recognised that extreme circumstances could compel a person to commit a crime, and it even categorised such circumstances as a form of compulsion,[99] but it was nevertheless unwilling to exculpate such actors.

---

[94] Macdonald, (1st edn, 1867), at p.17.
[95] *Ibid*.
[96] There are semantical differences, but nothing of substance. The fourth and fifth editions of the work were published posthumously without Macdonald's involvement.
[97] Anderson, p.6.
[98] *Cochrane v HM Advocate* 2001 SCCR 655 at 663.
[99] Hume described such extreme circumstances as "compulsion by want" and "pressure of extreme want": i, 54; Anderson curiously treats necessity as a separate consideration in the first edition of his book, but by the second edition he treats it as a type of compulsion: *op cit*. (1st edn, 1892), p.6-7, (2nd edn, 1904), p.16; and

Anderson's account is worth highlighting for three reasons. First, he is the only author to claim that necessity was "a good defence to a criminal charge", recognising it as a good answer to charges of piracy or involvement in a rebellion.[100] Far from suggesting that Anderson was alone in recognising necessity as a *sui generis* concept in Scots law, it more likely reveals that the other authors of this time were treating such scenarios as instances of compulsion, because we know from their accounts that piracy and being caught up in a mob/rebellion were good grounds for such a defence.

The second point is related to the first and concerns the fact that, for whatever reason, Anderson decided between the first and second editions of his text that necessity was better thought of as a type of compulsion, and thus moved the concept from its own separate category into his treatment of compulsion.[101] Again, this gives us good reason to think that, during the nineteenth century, and insofar as a defence of necessity was recognised, necessitous situations were being understood by the legal system as a variation of the more established compulsion defence.

Finally, Anderson's account is interesting because it is unique in relying on the English authority of *Dudley and Stephens*[102] to support the proposition that necessity is a good plea in Scots law.[103] From the outset, this proposition must be questioned given that the case itself casts heavy doubt on the existence of the plea in English law, never mind Scots law. This now infamous case involved survival cannibalism in a maritime setting – a shipwrecked crew killed and ate their poorly cabin boy in order to stave off starvation and improve their chances of rescue. They were, in fact, rescued four days later and thereafter charged with the young boy's murder. Rejecting the defendants' plea of necessity as a defence, the men were convicted of murder, only to have their sentences commuted to six months imprisonment by executive decision. They were labelled murderers by the courts but were heroes in the eyes of the community.[104]

The case is legally fascinating for many reasons, both substantive and procedural, and it is thought that the decision to deny the defence was a political one, viewed as an opportunity to make a condemnatory statement about what had emerged as a standard and

---

finally the fifth edition of Macdonald is the first to make any reference to 'extreme want', and again classifies it as a type of compulsion: (5th edn, 1948), p.11.

[100] Anderson, *op cit*. (1st edn, 1892), p.7.

[101] Anderson, *supra* fn.99.

[102] (1884) 14 QBD 273.

[103] The case only appears in the fifth and final edition of Macdonald, long after the author's death.

[104] AWB Simpson, *Cannibalism and the Common Law: The Story of the Tragic Last Voyage of the Mignonette and the Strange Legal Proceedings to Which it Gave Rise* (Harmondsworth: Penguin, 1986), at pp.80 & 83.

accepted practice of survival cannibalism in the seafaring trade.[105] However, owing in large part to a bizarre sequence of procedural events orchestrated to secure conviction,[106] the case did very little to advance the legal knowledge of how jurisdictions should treat actors who respond to situations of extreme pressure generally, except to say that one should rather die than take a life to save one's own.[107] Further, Lord Coleridge, giving the judgment of the court, was at pains to point out that necessity was a questionable plea in English law, and suggested that it was not even a good defence to charges of theft to stave off hunger.[108] In this sense, the case actually did more to set back the defence than further its understanding, and the result was that it became unclear if any defence of necessity existed at all.

Anderson's inclusion of *Dudley and Stephens* is therefore curious because it suggests a different understanding of the *ratio* of that case which holds broader implications for the role of English law in establishing a necessity defence in Scots law. Specifically, it suggests that Anderson regarded the case as authority for the existence of a necessity defence, with the limitation that it could not be pled as a defence to murder (the alternative being that he had misunderstood the case entirely). It therefore also suggests that the English jurisprudence on the defence was to be regarded as influential on, and indeed perhaps even formative of, the Scottish understanding of necessity. It is therefore worth examining the English understanding of necessity and compulsion in more detail to see how it might have influenced its evolution in Scotland.

2.4.2 The influence of English law

There was certainly a greater (documented) awareness in England than in Scotland of what necessity was and how it differed from other similar concepts, such as self-defence. As early as 1551 a principle of necessity was recognised where the court stated that "a man may break the words of the law, and yet not break the law itself… where the

---

[105] Simpson, *op cit*. at p.200.
[106] The trial judge Baron Huddeston feared that, given the men's popularity with the general population, allowing a jury to decide the case would result in an acquittal. He therefore set up the case for the jury to provide a 'special verdict' (as opposed to the general verdict normally given), an archaic system whereby the jury would be asked to find which facts pertained, and then a bench of judges would be asked to decide the legal conclusion based on those facts: *ibid* at 209. Forced into the situation of determining the legal question of whether the killing was necessary or not, counsel for defence Arthur Collins was thus precluded from arguing that necessity could partially exonerate and that a lesser charge, such as manslaughter, should be substituted: *ibid* at 229.
[107] (1884) 14 QBD 273 at 283 and 287 per Lord Coleridge.
[108] *Ibid* at 283.

words of them are broken to avoid greater inconvenience, or through necessity, or by compulsion".[109] Arnolds and Garland highlight the case's reference to the New Testament example of eating sacred bread or taking another's corn out of hunger.[110] Glazebrook suggests that Francis Bacon, publishing in 1630, was the first English writer to distinguish between killing in self-defence and killing to avert a threat to life for which the victim was not responsible, classifying the latter cases as examples of necessity.[111] A defence was granted on the basis that

> "The law chargeth no man with default where the act is compulsory and not voluntary, and where there is not a consent and election; and, therefore, if either there has been an impossibility for a man to do otherwise, or so great a perturbation of the judgement and reason as in presumption of law man's nature cannot overcome, such necessity carrieth a privilege in itself."[112]

Thus, and following the earlier court decision, where a man stole to "satisfy his present hunger", Bacon considered there would be no felony or larceny.[113] But neither Hale nor Blackstone agreed with this assessment, the former stating that despite such circumstances the theft would still be a felony,[114] and both cautioned against the political issues that could arise by such a defence rendering property insecure. This view, essentially mirroring Hume's views on *burthynsack*, had prevailed leading into *Dudley and Stephens*, and indeed has done ever since.

An examination of Hale's treatment of necessity also lends credibility to the claim made above suggesting that Scottish jurists treated necessity as a form of compulsion. Hale begins his discussion with the claim that "all compulfion carry with it fomewhat of neceffity".[115] He goes on to state that, in contrast, not all kinds of necessity originate from an external compulsion or force.[116] Compulsion, then, is at least in part a type of necessity.

---

[109] *R v Fagossa* [1551] 1 Plowd. 1; 75 Eng. Rep. 1.
[110] EB Arnolds & NF Garland, "The Defence of Necessity in Criminal Law: The Right to Choose the Lesser Evil" (1974) 65 *Journal of Criminal Law & Criminology* 289 at 291, citing *Matthew* 12:3-4.
[111] Glazebrook, *op cit*. at 110.
[112] F Bacon, *The Elements of the Common Lawes of England*, Volume 13 (London: Assignes of I. More Esq, 1630), p.131.
[113] *Ibid*, p.160.
[114] M Hale, *History of the Pleas of the Crown*, Volume 1, (London, 1736), pp.53-54, para [54]. Hale's *Pleas* was directed by the House of Commons to be printed in 1680 but this did not happen until 1736: see *Pleas*, front matter titled 'Extract from the Journal of the House of Commons', dated 29 November 1680. For Blackstone, see *op cit.* at pp.31-2. Although note the contradictory comments made at p.15 in relation to extenuating conditions in the previous chapter on the nature of crime and punishment: "…theft, in a case of hunger, is far more worthy of compassion, than when committed through avarice or to supply one in luxurious excesses".
[115] Hale, *Ibid* at p.53, para [54].
[116] *Ibid*.

Likewise, Blackstone speaks of *duress per minas*, i.e. "threats and menaces, which induce a fear of death or other bodily harm", as being another "species of compulfion or neceffity".[117] Hale's views, taken in totality, represent the idea that compulsion represented the kinds of necessity that the law was willing to admit. Beyond compulsion, Hale points to the killing of thieves or rioters in the process of attempting to maintain the peace by public officials as being a justified necessity. Other kinds of necessity, such as of extreme want found in *burthynsack* and survival cannibalism, thus represented a step too far, because they threatened to "let loose, as much as they can… all the ligaments of property and civil society".[118] This discussion appears to mirror the justifications Hume later gave for why compulsion by extreme want should not be admitted as a defence in Scots law.[119] There therefore appears to have been a common understanding, between both jurisdictions, that compulsion and actions taken to preserve the peace represented the only acceptable forms of necessity.

Blackstone also recognised a lesser evils form of necessity in his *Commentaries* in similar terms to Hale, understood as a form of defect of will and to be distinguished from compulsion by force or fear on the basis that necessity results from "reason and reflection, which act upon and constrain a man's will, and oblige him to do an action".[120] Blackstone argued that the will in such circumstances should be treated as passive – if it were active it was only so in the rejection of the greater evil.[121] Necessity should thus, despite involving the defendant's 'reason and reflection', be seen as a defect of will.[122]

Both Hale and Blackstone's understanding of necessity appears to have been heavily influenced by their views on the importance of maintaining both the primacy of the law and civil order. We saw above their reluctance to give a defence to the starving thief and, like Hale, Blackstone gives as good examples of necessity any force used when undertaking a citizen's arrest of a person committing a capital offence, or when dispersing a riot, and suggests that it may be both justifiable and necessary to kill the offenders, rather than permit a murderer to escape, or the riot to continue.[123] It appears Blackstone recognised that much criminal activity could be traced back to social conditions, but nevertheless considered that the potential danger to public order necessarily precluded

---

[117] Blackstone, *op cit*. at p.30.
[118] Hale, *op cit*. at p.54, para [54].
[119] Hume, i, 54-55.
[120] Blackstone, *op cit*. at p.30.
[121] *Ibid* at p.31.
[122] Despite the justificatory overtones: on this point see A Wertheimer, *Coercion* (Princeton University Press, 1987) at p.147.
[123] Blackstone, *op cit*. at p.31.

these conditions from grounding a defence.[124] Thus, necessity appears to have only been recognised when it was not in direct contention with the primacy of the law – in other words it seems to have operated in a way analogous to the doctrine of double-effect, whereby the accused could point to some aspect of law or civil order that they were upholding in committing their offence. If none could be found, no defence existed.

2.4.3 Summary

The argument put forward here then is that necessity likely *was* recognised in early Scots law, insofar as the Scottish legal system understood compulsion as its predominant valid form.[125] The two were one and the same, and it was not until the twentieth century that a new, separate form of necessity emerged in both Scots and English law,[126] creating a more pronounced distinction between offences committed under compliance to avoid a danger to life, and those committed 'voluntarily' to the same end. In this sense, insofar as necessity as we understand the defence in modern terminology was previously recognised in Scots law, it was likely only in relation to offences which had as their goal some aspect of maintaining the peace or upholding the law. Personal struggles, no matter how dire, were insufficient as they posed a direct threat to the primacy of the law.

Nevertheless, this distinction had always been there, because it is clear from the historical accounts outlined above that if compliance was present, then a defence was more likely. This may say something about the importance of *dolus* or an *animo furandi* in each legal system – in cases where there was another blameworthy human agent, seen to be orchestrating the situation, it would be easier to accept that the accused lacked the necessary malice because justice could still be seen to be done (to the coercer). In any case, even the rules on compulsion demonstrated a lack of conceptual coherence, based on an unwillingness by the law to extend its exculpatory effect beyond extremely dire circumstances, such as civil unrest. The overwhelming concern has always been with maintaining the integrity and efficacy of the criminal law, even in situations where liability

---

[124] On the contradictory nature of Blackstone's views, see TA Green, *Verdict According to Conscience* (University of Chicago Press, 1985) at pp.295-6.

[125] Other examples did of course exist – namely a defence for piracy and for homicide when attempting to maintain the peace.

[126] For English law, see: *R v Willer* (1986) 83 Cr App R 225; *R v Conway* [1989] QB 290; and *R v Martin* (1989) 88 Cr App R 343 establishing a new defence of duress of circumstances (discussed below at section 3.3). For Scots law, see *Moss v Howdle* 1997 JC 123 establishing the defence of necessity. Each of these cases is discussed in more detail below.

was thought to be unfair. This fact is, of course, most obviously evidenced by the case of *Dudley and Stephens*.

2.5 Conclusion

The defences of necessity and coercion have experienced a rather icy reception throughout Scottish legal history. An unwillingness to compromise or dilute the rule of law for fear of a potentially negative impact on civil order has meant that these defences emerged in the twentieth century conceptually underdeveloped. In contrast, other reactive defences such as self-defence,[127] and even partial forms of exculpation such as provocation,[128] had been recognised more or less in their modern forms as early as the seventeenth century. This perhaps demonstrates how conceptually difficult the ideas of compulsion and necessity were to reconcile with the ideals of the rule of law and maintaining civil order. Self-defence was easy to explain in these terms and, as a result, was warmly received and debated in early Scots criminal law and entered the twentieth century fully fleshed out as a long-standing substantive defence. Even provocation could be understood in these terms by refusing the defence in cases of verbal provocation,[129] such that it could be said to ensure civil order was maintained as far as possible. In contrast, and despite the brief rise to fame precipitated by *Dudley and Stephens*, both necessity and coercion would again lay dormant until the 1970s when we finally started to see accused persons attempting to rely on these defences against charges for less serious offences.

In the next two chapters, this thesis examines these cases and the resulting precedent which forms the current legal rules on the necessity and coercion defences in Scots law. This analysis includes a preliminary examination of how both defences came to be governed by the same rules despite Hume's clear reluctance to acknowledge any kind of general necessity principle, as well as an overview of the fractured English conception of necessity which has continued to influence the approach of the Scottish judiciary. I shall undertake this literature review with the goal of appropriately framing the defences and

---

[127] Mackenzie, MC, Title 11, 2-5; Bayne, p.84-86; Forbes, Part I, Book IV, Chapter I, 3, p.110.
[128] Mackenzie, MC, Title 11, 3 & 14 (he argues that killing to defend one's honour is a form of revenge and not self-defence, and thus should be rejected, but states that a father or husband may kill their daughter or wife if caught committing adultery); Bayne, p.81 (stating that adulterers are a class of person that may be killed with impunity). Cf. Forbes, Part I, Book IV, Chapter I, 1, p.102 discussing culpable homicide via provocation. Given that adultery was a criminal offence at this time, it was perhaps easier to establish a defence on a basis analogous to self-defence and the prior 'wrong' of the victim reducing culpability.
[129] See, e.g. *ibid*; Hume, i, 226.

their current parameters, in order to highlight the issues which create the need to consider an alternative framework for understanding reactive defences like necessity and coercion. Such a framework will be considered in part two of this thesis.

3. Contemporary Scots law on Necessity and Coercion: A Danger of Death or Serious Injury

3.1 Introduction

As mentioned in the previous chapter, it would take almost a century from the decision in *Dudley and Stephens* before necessity and coercion would be discussed in the Scottish courts in any capacity. Indeed, even in the English courts from which *Dudley* arose cases were sparse,[1] despite some commentators claiming that a defence of necessity was well recognised.[2] Gordon suggests that this lack of formal recognition sometimes led to harsh results, with the accused in *Watson v Hamilton*[3] being refused a defence to driving under the influence after taking a pregnant woman to the hospital when she started bleeding in the middle of the night. It was after midnight and none of the nearby payphones were working, so the accused drove her to the hospital and was charged, convicted and subsequently disqualified from driving.[4] Only in 1997 did the Scottish appeal court formally recognise necessity.[5]

The first modern Scottish appeal court decision which recognised either defence was *Thomson v HM Advocate* in 1983,[6] a case concerning a coercion plea in defence to charges of armed robbery. There, Lord Justice-Clerk Wheatley confirmed the existence of coercion as a valid defence in Scots law and outlined the requirements to be satisfied before it could be put to a jury. At trial, Lord Hunter cited the four requirements set out by Hume: an immediate danger of death or serious injury; an inability to resist the violence; a backwards and inferior role in the perpetration; and a disclosure of the fact, as well as

---

[1] The first cases to offer any substantive discussion of necessity after *Dudley and Stephens* were *Evans v Wright* [1964] 1 WLUK 183 and *R v Baines* (1970) 54 Cr. App. R. 481; the plea was unsuccessful in both. The earliest case discussing duress is *R v Steane* [1947] 1 All ER 813 (although the case was decided on a strenuous interpretation of the intention requirement).

[2] GL Williams, *Criminal Law: The General Part* (1st edn, London: Stevens & Sons; 1953) at p.216.

[3] 1988 SLT 316.

[4] This result was particularly harsh because the accused was also unsuccessful in challenging his disqualification, which is merely a reduction of penalty rather than full exoneration. Discussed in Gordon, *Criminal Law*, at para. 13.21. See also *Morrison v Valentine* 1991 SLT 413. Cf. the earlier sheriff court decision of *Tudhope v Grubb* 1983 SCCR 350 where necessity was successfully pled, discussed below. Many of the early cases which considered necessity before its formal recognition did so in the context of driving offences, and concerned whether the necessity of the situation ('special reasons') should prevent the judge from disqualifying the accused: see, e.g., *Watson v Hamilton* 1988 SLT 316; *McLeod v MacDougall* 1989 SLT 151; *Hamilton v Neizer* 1993 JC 63.

[5] *Moss v Howdle* 1997 JC 123, discussed below.

[6] 1983 JC 69.

restitution of the spoil, on the first safe and convenient occasion.[7] This statement was approved by the appeal court and has since become the foundation for the substantive defences of coercion *and* necessity in Scots law.

The following two chapters shall outline the current rules which apply to both necessity and coercion in Scots law based on the passage found in Hume and other, additional requirements which have emerged in the modern era. This chapter shall begin by explaining some preliminary matters which will help to frame the rest of this expansive literature review. Specifically, we begin with an explanation as to why Hume's influential passage on compulsion can be said to apply to both coercion and necessity. This is followed by a brief history of the necessity defence and its variations in English law, including the 'duress of circumstances' formulation pioneered by the Court of Appeal in the eighties, which has had a profound impact on the development of a necessity defence in Scotland. Finally, this chapter shall undertake a detailed exposition of Hume's most important requirement – that for an accused to have a defence there must have been an immediate threat of death or serious injury. This analysis will centre around providing a deeper understanding of the terms 'threat' and 'danger', including issues such as the nature of implied threats, to whom a threat must be directed, and issues pertaining to the quality of the accused's unlawful action, such as its prospects of removing the danger and its proportionality to the averted harm.

By outlining the defences in this way over the next two chapters, this thesis aims to provide the reader with a detailed exposition of this area of law both for ease of reference, and to better frame the problems that arise in relation to these defences in practice, providing the impetus for an alternative understanding of these defences which shall be developed in part II of this thesis. In addition, however, it is hoped that the following analysis presents a novel way of approaching the discussion of these defences. Despite the courts consistently characterising these defences as being almost identical and subject to the same rules, typically each defence is given individual treatment in the academic literature. By considering the defences in tandem, I intend to provide a more coherent understanding of each defence and its current place within Scots criminal law.

[7] *Ibid* at 75. See also section 2.3.2 above.

3.2 Preliminary Issues: Adapting Hume for Modern Scots law

The forgoing exposition of necessity and coercion in Scots law is based on a particular analysis of the historical sources, such that this thesis treats sources for each defence as applicable to both (unless expressly stated otherwise). In this preliminary section I shall therefore justify this reasoning, focusing on how Hume's treatment of compulsion has been adapted to become foundational to both necessity and coercion.

3.2.1 Hume as the foundation for both modern defences

First, it should be explained why the judgment in *Thomson* (and thus Hume's requirements) can be taken to provide the foundations for both coercion and necessity in Scots law. Certainly, the court in *Thomson* had no such grand aspirations, but in the subsequent case of *Moss v Howdle*[8] the court held that both defences were available as substantive defences where an accused acted under an immediate danger of death or great bodily harm.[9] To reach this conclusion, Lord Justice-General Rodger drew on the analogous situation in English law, claiming that the 'necessity' defence understood in Scots law was to be regarded as equivalent to the recently established defence of 'duress of circumstances' which operated as part of the broader duress defence.[10] Thus, just as duress had been expanded to include both human threats and other circumstances, Lord Rodger held that if a defence of coercion was available to an accused who committed a crime in order to escape a threat from a third party, then there was "no reason why [the defence] should be excluded simply because the immediate threat of death or great bodily harm which the accused is trying to evade arises from, say, a natural disaster or from illness".[11]

Unfortunately, the reasoning of the court is incoherent, which results in the judgment being unclear on the broader legal rules which apply to necessity. From the outset, Lord Rodger makes it clear that both defences will be available where there is a threat of death or serious bodily harm but, beyond this undoubtedly important point, we are otherwise left to decipher precisely what the judgment means for the broader context of the necessity defence. Indeed, Lord Rodger begins his exposition by rejecting counsel's

---

[8] 1997 JC 123.
[9] Although this was not the first case to draw such a conclusion. In *McNab v Guild* 1989 JC 72 at 75 the court tentatively agreed that, if there were to be a defence of necessity in Scots law, it should only be applicable on the same basis as coercion: i.e. in response to threats of death or serious injury.
[10] 1997 JC 123 at 127. See also *supra*, chapter two, fn. 126 above. 'Duress of circumstances' will be discussed in more detail below.
[11] *Ibid*.

argument that situations of necessity were covered by Hume's passage on compulsion, stating that this would take the scope of the defence "further than is envisaged in the passage of Hume".[12] However, after instead aligning necessity with the English defence of duress of circumstances, his Lordship cites with approval a passage from *R v Howe* where Lord Hailsham claimed that the distinction between necessity and coercion was one "without a relevant difference, since on this view duress is only that species of the genus of necessity which is caused by wrongful threats".[13] This point is iterated several times, including a statement that "it makes no difference to the possible availability of any defence that the danger arises from some contingency such as a natural disaster or illness rather than from the deliberate threats of another" and that "the law should regard all of these threats in the same way".[14]

If the nature of the threat is irrelevant, it is difficult to see on what basis necessity is being distinguished from coercion, such that an extension of Hume's provisions to necessity is inappropriate. A potential counter point may be to draw attention to the fact that Lord Rodger had stressed that the passage from Hume "does not purport to give a full description of the defence as it is to be applied in all circumstances", and thus his Lordship's subsequent description of necessity should be treated as an extension of sorts. It seems clear, however, that his Lordship intended for the rules to apply equally to both defences. The implication is therefore that Hume's passage on compulsion is to be read as applying equally to both coercion and necessity defences in Scots law, albeit not as a "full description of the defence" or any absolute statement of the rules.[15] However, given the historical context discussed above in the previous chapter which saw that a very limited version of necessity was in fact included in the definition of compulsion, this interpretation of Lord Rodger's judgment is perhaps not quite the bastardisation of the concept that it initially appears to be.

In any case, *Moss v Howdle* is often cited for the proposition that the rules applying to coercion should apply equally to the defence of necessity in Scots law but, in reality, Lord Justice-General Rodger's argument goes a step further: it suggests that Hume's passage is foundational to both. Despite this, Hume's compulsion test is very rarely (if ever) cited in relation to cases of necessity or in textbooks discussing the topic, being exclusively discussed in relation to coercion. Indeed, in the subsequent case of *Lord*

---

[12] *Ibid*.
[13] *Ibid* at 128, citing Lord Hailsham in *R v Howe* [1987] AC 417 at 429.
[14] *Ibid*.
[15] *Ibid*.

Advocate's Reference (No 1 of 2000), where the court reaffirmed the immediacy requirement in cases of necessity, no reference was made to the passage by Hume despite the court affirming the *ratio* of *Moss v Howdle*.[16] Of course, by affirming *Moss*, the court in *Lord Advocate's Reference* can be taken to implicitly agree that the passage in Hume forms the basis of the defence: working backwards the court in *Lord Advocate's Reference* affirmed that "[t]he law of Scotland is as declared in *Moss v Howdle*"[17] and the court in *Moss* based their test for necessity on Hume's test of compulsion. Thus, Hume's test should be regarded as the infrastructure for both defences. In the following sections, I shall therefore outline each of Hume's requirements in turn, in the context of both coercion and necessity, followed by sections detailing the other requirements which have emerged in the modern era, which are equally applicable to both defences.[18]

3.2.2 The severity of the crime

One important qualification which is discussed in *Thomson* and reiterated in *Moss v Howdle* (i.e. in relation to both defences) is the fact that Hume's requirements were not intended to be comprehensive. There was some discussion, and indeed confusion, about precisely when the requirements are supposed to be activated – this can be attributed to the statement made by the court in *Thomson* that "A defence of coercion is recognised in the law of Scotland… Hume restricts it to 'atrocious crimes'".[19] This statement appears to be an outlier in the context of the case, as well as inconsistent with the passage in Hume itself,[20] and thus it is submitted that the phrase is better understood as an error and read out of the judgment. Excluding this difficult phrase, *Thomson* can be taken to state that, whatever the rules are for coercion generally, when the offence is serious, to be determined by the nature of the crime and its attendant circumstances,[21] Hume's requirements will apply. For their part, the court in *Moss v Howdle* highlighted this peculiar comment made in *Thomson*, reaching a similar conclusion by pointing out that it would be an "odd legal

---

[16] *Lord Advocate's Reference (No 1 of 2000)* 2001 JC 143 at [34] (approving *Moss v Howdle*) & [37] (reaffirming immediacy requirement).
[17] *Ibid* at [55].
[18] See, e.g. *Cochrane v HM Advocate* 2001 SCCR 655 at 666.
[19] 1983 JC 69 at 78.
[20] Hume, i, 52 (emphasis added): "this is at least a somewhat difficult plea, and can hardly be serviceable in the case of a trial for any atrocious crime, *unless it have the support of these qualifications*".
[21] 1983 JC 69 at 78.

system indeed" which allowed coercion to extinguish guilt for armed robbery but not for speeding offences.[22]

However, it is in the application of Hume's requirements to the circumstances in *Moss v Howdle*, i.e. an objectively 'non-serious' offence, which then brings everything full circle and effectively means that the requirements should be taken to apply in *all* putative cases of necessity.[23] It is not entirely clear whether the courts would extend this logic to cases of coercion for non-serious offences today, particularly because Lord Justice-General Rodger does not appear to have been aware that his comments would have such far-reaching consequences. He was of course talking in the context of the necessity defence specifically, but it is difficult to distinguish the two defences when the premise for his judgment was based on their amalgamation. However, at the time of writing there have been no Scottish cases concerning a coercion defence being pled to a charge for a non-serious offence, so the issue is still very much a live one.

One other possible interpretation of Hume's requirements in the context of their application to atrocious crimes is to read the passage as stating that, in general and for there to be a substantive defence, the requirements must be met but, where they are not met, circumstances of compulsion not in themselves sufficient to ground a complete defence may nevertheless operate to reduce liability by way of mitigation of sentence. This interpretation has the advantage of mapping to the limited case law on the topic.[24] Both *Trotter*[25] and *Cochrane*[26] proceeded on the basis that an immediate danger which the accused was unable to resist or avoid was essential, and that future threats would be insufficient to ground the plea of coercion. Indeed, there are no examples of coercion being successfully pled where this requirement was not satisfied. The same position has been taken with respect to necessity.[27] The proposed interpretation would also establish Hume's test as the general basis for necessity and coercion in all cases, rather than leaving an

---

[22] 1997 JC 123 at 126.
[23] Although cf. *R v Conway* [1989] QB 290 at 298 per Lord Justice Woolf: "No wider defence to reckless driving is recognised. Bearing in mind that reckless driving can kill, we cannot accept that Parliament intended otherwise." Hypothetically, any driving offence has the potential to kill given the inherent nature of driving and the dangers involved, but it seems unlikely that this logic would be extended to minor driving offences, such as speeding as in *Moss*.
[24] See, e.g., Lord Macfadyen's observations in *Cosgrove v HM Advocate* 2008 JC 102 at 106: "The test to be applied by the jury in reaching a view on the defence of coercion is a strict one, but a different, less stringent, test falls in our view to be applied to whether the fact that the appellant was acting under pressure short of legal coercion may be relevant mitigation in assessing his culpability with a view to fixing the punishment part of his life sentence."
[25] *Trotter v HM Advocate* 2000 SCCR 968.
[26] *Cochrane v HM Advocate* 2001 SCCR 655.
[27] See, e.g. *D v Donnelly* 2009 SLT 476 at 477; *SB v HM Advocate* 2015 JC 289 at 297.

apparent gap where different, unknown rules apply for any crime which is deemed not to be 'serious'. It is submitted that this would provide clarity to this area of law, as it is currently hard to imagine just what sort of test would be permissible in cases of a non-serious nature.

It might be suggested that the above argument involves rejecting the statements made in *Thomson* and *Moss* to the effect that Hume was not intending to "lay down any absolute rule"[28] or provide "the entire law of Scotland on the topic".[29] To this reservation I would remind the reader that the "topic" Hume was discussing was, in fact, 'compulsion' generally, and not the substantive defences of coercion and necessity. Compulsion is a much broader topic, if not in substance, then certainly with respect to procedure. As understood by Hume, compulsion was a plea that could completely excuse an accused, but might also only provide grounds for mitigation.[30] With this in mind, the proposal to apply Hume's requirements to all instances of coercion or necessity pled as a substantive defence irrespective of the offence charged is not inconsistent with his statement that his treatment of compulsion was not absolute. In any case, the rules received by Hume have very much become 'absolute' in practice.

3.2.3 The relevance of a 'well-regulated society'

Indeed, the above construction also allows us to distance ourselves from another unhelpful aspect of the necessity and coercion defences established by Lord Wheatley's comments in *Thomson*, which outlines the importance of the presence of the ordinary conditions of a well-regulated society to the availability of the plea.[31] His Lordship suggests that the requirements would be necessary in "the type of society in which we live today" because where there is "time and opportunity to seek and obtain the shield of the law… then recourse should be made to it".[32] In other words, if there was no immediate

---

[28] 1983 JC 69 at 77.
[29] 1997 JC 123 at 127.
[30] Hume, i, 50 (my emphasis added): "A person is not therefore guilty of treason, who being in a part of the country that is commanded by the rebels, yields them against his will supply of money, or arms and provisions… *Perhaps this may even go the length of excusing him*, though he be with them for a time in arms, provided he prove an outrageous compulsion in that particular, and that he quit their service on the first opportunity, and do no act of hostility while he remains with them, that is not fairly imputable to the constraint of his situation." See also at i, 55, in relation to the form of necessity found in *Regiam Majestatum* known as "burthynsack", which allows a man to go unpunished for the theft of meat, if he is compelled to do so by hunger. Hume, dismissive of the concept, argued this was not a complete defence but rather a plea in mitigation.
[31] 1983 JC 69 at 76-7.
[32] 1983 JC 69 at 77.

danger, there was no logical reason for an accused not to rely on recourse to legal methods of response. This statement was, however, somewhat tempered by his follow up where he admitted that,

> "In saying this we are conscious that even in the ordinary condition of a well-regulated society there may be circumstances where a person is exposed to a threat of violence to himself… from which he cannot be protected by the forces of law and order and which he is not in a position to resist. If such a situation arose it would have to be determined on its facts".[33]

It is not immediately clear how this qualifies Lord Wheatley's initial statement. On the one hand, it could be taken to suggest that, in special circumstances, a defence of coercion will be available even where there is no immediate danger. Alternatively, it could mean only that the court was aware of and acknowledged the serious pressure and distress that persons under duress experience, and that such circumstances may be utilised to mitigate punishment. Given that the context of the statement was to qualify a previous sentence in which the court endorsed Hume's requirements as representing the law of Scotland in relation to coercion as a complete defence, the first interpretation seems preferrable.

However, Lord Wheatley's comments on the relevance of a well-regulated society are problematic for two reasons. First and foremost, they are, in a sense, contradictory. He seems to suggest that only in the absence of a well-regulated society will Hume's requirements not apply, but then adds that there may be some circumstances, undefined, in which the requirements will not apply despite a well-regulated society being present. This kind of mystification of the rules surrounding a substantive defence, indeed arguably two substantive defences, is unhelpful. This leads into point two which is that the contradiction invites uncertainty into the law, and this should be avoided. There is a distinction to be drawn between trying to keep the defences 'within narrow bounds' on the one hand and being overly obtuse with the requirements and their application, on the other. It is respectfully argued that qualifying the rules on coercion claims by reference to the prior existence of civil order is, in light of the other requirements, superfluous and unhelpful.

---

[33] *Ibid* at 78.

3.3 English Necessity and 'Duress of Circumstances'

Before moving on to outline the various requirements for necessity and coercion in contemporary Scots law, there remains one final issue to be discussed in relation to these defences and their relationship with the English law on the subject. As we saw in the previous chapter, the English jurisprudence has been historically influential in the development of necessity in Scots law. This remains true today, and thus an understanding of the current English law on the topic is appropriate to fully appreciate the reasoning behind some of the decisions which affect Scots law. In fact, there is good reason to believe that the understanding of necessity received into Scots law in *Moss v Howdle* was exactly the type of necessity first developed in England in the 1980s known as 'duress of circumstances', to be contrasted with the English equivalent of coercion, duress by threats.[34] In general, I shall highlight important points of English law as they arise in relation to Scots law in the sections which follow, but the concept of 'duress of circumstances' should be explained at the outset, in terms of its relationship to the concepts of necessity in both English and Scots law, to avoid confusion going forward.

3.3.1 Evolution of duress of circumstances

As discussed in the previous chapter, necessity as a defence is recognised in English law by *Dudley and Stephens*, but to what extent it provides exculpation is unclear. The reader will recall that in *Dudley and Stephens* the court held that necessity was not a good defence to a charge of murder, although this says nothing about its availability in response to other charges.[35] Nevertheless, a clear reluctance[36] to expand the kind of necessity espoused in *Dudley and Stephens*, one which is characterised as a justification and based on competing values, has encouraged the English courts to develop a separate defence, one borne out of the related defence of duress by threats – a defence which is understood as a concession to human frailty (i.e. wrongful but not blameful conduct), and thus follows strict requirements. On this basis, the appeal court in *R v Willer*[37] rejected the availability of a necessity defence to charges of reckless driving but held that a defence of duress could be pled where someone was "driven by force of circumstance into doing what

[34] Known more colloquially as 'duress'. It can also be referred to as duress *per minas*: see, e.g. Lord Kilbrandon, "Duress *Per Minas* as a Defence to Crime: I" (1982) 1(2) *Law and Philosophy* 185.
[35] See section 2.4.1 above.
[36] *R v Denton* (1987) 85 Cr App R 246, particularly at 248 per Caulfield J doubting whether necessity as a defence could be pled successfully against charges of reckless driving.
[37] (1986) 83 Cr App R 225.

he did".[38] The defendant had been charged with reckless driving after mounting the pavement to avoid a gang of youths who were intent on doing the accused and his passengers harm. There was therefore no command that the accused followed – Willer chose to mount the pavement of his own volition.

This line of authority was continued in *R v Conway*,[39] where the court discussed *Willer* in detail. Lord Woolf, rather than reject necessity outright as in *Willer*, held that necessity could be pled as a defence to charges of reckless driving but only where the facts "establish 'duress of circumstances'".[40] Thus, rather than create a distinction between duress of circumstances and necessity, the court stated that duress is an example of necessity and that the name "does not matter"; what matters is that, whichever name the defence may take, it is "subject to the same limitations as the 'do this or else' species of duress".[41] In other words, Lord Woolf held that the same rules which apply to duress by threats should apply to any potential defence of necessity – whether known by that name or another – if it were to be accepted in English law as a substantive defence. In that sense, *Conway* is to English law what *Moss* is to Scots law in terms of attempting to unify the rules applicable to the two defences,[42] albeit the comparison is not so neat given that various forms of necessity can be said to exist in English law (discussed below). Lord Woolf's confirmed that the duress of circumstances defence would only be available if the defendant could be said to be acting to avoid a threat of death or serious injury from an *objective* standpoint.[43] Thus, it would be insufficient if the accused honestly but irrationally believed that there was such a threat – the belief also had to be reasonable for duress of circumstances to be applicable.

Following closely from *Conway*, and solidifying duress of circumstances as a novel defence in English law, the court again considered the extent to which 'necessity' might operate as a defence in *R v Martin*,[44] another case involving a driving offence. Martin drove while disqualified after his suicidal wife threatened to harm herself if he did not drive at the requested time.[45] At trial Martin's claim of necessity was rejected on the basis

---

[38] *Ibid* at 227.
[39] [1989] QB 290.
[40] *Ibid* at 297.
[41] *Ibid*.
[42] W Chan and AP Simester claim that the criteria for duress by threats and duress of circumstances "are in essence the same": "Duress, Necessity: How Many Defences? (2005) 16 *King's Law Journal* 121 at 121.
[43] [1989] QB 290 at 298.
[44] (1989) 88 Cr App R 343. The judgment refers to a defence of 'necessity' but, when discussing the judicial history, states that *Conway* establishes the defence of 'duress of circumstances': at 345-6.
[45] It is curious that the defence of duress *by threats* was not considered as more appropriate, a fact noted by the court in *R v Cole* [1994] Crim LR 582 at 583.

that the offence was one of strict liability, but on appeal Simon Brown J, giving the judgment of the court, relied on the decision in *Conway* as a foundation for holding that a defence of necessity was recognised in "extreme circumstances" which he outlined.[46] In order for the defence to succeed there had to be: i) pressure upon the accused's will, arising from the wrongful threats or violence of another, or from other objective dangers; ii) a reasonable and proportionate response by the accused, to be assessed objectively; and, assuming the defence is open on the facts, iii) an affirmative response from the jury to the following two questions – a) was the accused impelled to act as they did because, as a result of what they reasonably believed to be the situation, they had good cause to fear that otherwise death or serious injury would result; and, if so, b) would a sober person of reasonable firmness, sharing the characteristics of the accused, have responded to the situation by acting as the accused did?[47]

After *Martin*, the scope of duress of circumstances was broadened such that it was available in more than just reckless driving cases. In *R v Pommell*,[48] for example, the Court of Appeal held that the defence should have been allowed to go to the jury where the defendant was convicted of possessing a firearm. After being found in bed with a loaded sub-machine gun, the defendant claimed that they had taken it from an acquaintance intending to use it and had planned on handing the gun over to his brother the next morning to give to the police. Unfortunately, the judgment invites confusion as it is given in the terms of lesser evils reasoning which provides the rationale for the justificatory version of the necessity defence, stating that it would be unsatisfactory to leave the situation where someone "commendably infringes a regulation in order to prevent another person from committing what everyone would accept as being a greater evil with a gun" to the discretion of the prosecuting authority not to prosecute, or to courts to grant an absolute discharge.[49] In *R v Abdul-Hussain*[50] the Court of Appeal held that the defence, whether by threats or from circumstances, was available to all substantive crimes except murder,

---

[46] *Ibid* at 345.
[47] *Ibid* at 346.
[48] [1995] 2 Cr App R 607.
[49] *Ibid* per Kennedy LJ at 613-4. See also Brooke LJ's judgment in *Re A (Children)* [2001] Fam 147 at 236: "I have described how in modern times Parliament has sometimes provided "necessity" defences in statutes and how the courts in developing the defence of duress of circumstances have sometimes equated it with the defence of necessity. They do not, however, cover exactly the same ground. In cases of pure necessity the actor's mind is not irresistibly overborne by external pressures. The claim is that his or her conduct was not harmful because on a choice of two evils the choice of avoiding the greater harm was justified."
[50] [1999] Crim LR 570.

attempted murder and some forms of treason.[51] That case involved charges of hijacking a plane, to which the defence was available.

3.3.2 Other variations of necessity in English law

The development of duress of circumstances has left the English law on necessity somewhat confusing and fractured.[52] There are presently, according to Stark,[53] at least four versions of necessity which exist in English law, assuming one takes as their starting point the broader necessity principle. Duress of circumstances is said to represent an/the[54] excuse variation of the defence, with other identifiable variations being 'Justificatory Necessity', self-defence and "best interests interventions".[55] Unlike Justificatory Necessity, self-defence requires the use of force which is then aimed against the threat itself,[56] and "best interests interventions" involve the harm doer seeking to further the interests of a person incapable of knowing what is in their best interests.[57] Other than *Dudley and Stephens*, English cases examining Justificatory Necessity in any detail are sparse. Stark discusses the more modern case of *Re A (Children)*,[58] where doctors applied to the court to allow them to carry out a separation of two conjoined twins via surgery which would save the life of the stronger twin (Jodie), but kill the weaker one (Mary). In the absence of such medical intervention, the strain of maintaining Mary would become overwhelming for Jodie and they would both die. The Court of Appeal considered that the surgery would constitute the killing of Mary, and thus would not be a lawful course of action unless an appropriate defence could be found.[59] In a lengthy judgment the court were unanimous in deciding that

---

[51] *Ibid*.
[52] E.g., treating necessity as synonymous with duress of circumstances: *R v Sasikaran Selvaratnam* [2006] EWCA Crim 1321 at para [32]: "the defence of necessity may be regarded as duress by force of circumstance".
[53] F Stark, "Necessity and *Nicklinson*" [2013] *Criminal Law Review* 949 at 950.
[54] Stark points out that the authorities point towards a takeover of the excusatory form of necessity by duress of circumstances, such that only the latter can be said to exist (or that they are synonymous), but the authorities are not conclusive: *ibid* at 952-3. For a recent example, see *R v Thacker* [2021] 1 Cr App R 21 at para. 92, per Lord Chief Justice (Burnett), where the idea that *R v Martin* (1989) 88 Cr App R 343 laid down rules for the necessity defence (as opposed to duress of circumstances) was undisputed.
[55] *Ibid* at 950. Although cf. *R v Jones* [2005] Crim LR 122 with commentary by DC Ormerod who notes, at 125, the courts' tendency to conflate duress of circumstances (excusatory form) with necessity (justificatory form).
[56] *Ibid* at 954-5.
[57] *Ibid* at 955. In that sense, Stark states that it is inappropriate to think of the person harmed as an innocent bystander who is in the wrong place at the wrong time. See, e.g. *Gillick v West Norfolk and Wisbech AHA* [1986] AC 112; *Re F (Mental Patient Sterilisation)* [1990] 2 AC 1.
[58] [2001] Fam 147.
[59] *Ibid* at 218, per Brooke LJ.

doctors could perform the surgery, but the three judges came to this conclusion on quite different grounds.

Ward LJ was unable to view Mary's existence as worthless, nor could he accept the logic that hastening her inevitable death was in her best interests,[60] but he considered that Mary posed a direct threat to the life of Jodie, and thus justified the operation on the grounds of self-defence.[61] Interestingly, he prefaced this conclusion by stating that he could see "no other way of dealing with [the case] than by choosing the lesser of the two evils and so finding the least detrimental alternative".[62] Walker LJ alluded to necessity being a potential ground for approving the elective surgery, stating that in the absence of parliamentary intervention the defence of necessity would have to develop on a case by case basis, and that he would extend it, insofar as an extension is necessary, to include the present case.[63] More generally, he argued that everybody has a right to bodily integrity and autonomy, and that there was therefore a strong presumption that an operation to separate them would be in their best interests, despite meaning certain death for Mary.[64]

Finally, Brooke LJ was the most explicit in holding that necessity *simpliciter* could justify the killing of Mary to save Jodie from certain death. After embarking on a lengthy exposition of the history of the necessity defence in English law,[65] he held that there were sound reasons to believe that a defence of necessity could operate to justify the operation, on the basis that the defence did not necessitate an emergency (in the normal sense of the word), nor was the threat required to constitute an 'unjust aggression' as there was a distinction to be drawn between necessity and private (self) defence.[66] In addition, and unlike *Dudley and Stephens*, the victim had already been self-designated for death, thus bypassing the difficult question of 'who is to judge this sort of necessity?'.[67] As per Sir James Stephen, Brooke LJ held that there were three requirements for the application of a necessity defence, which he considered satisfied. These were: i) the act is needed to avoid inevitable and irreparable evil; ii) no more should be done than is reasonably necessary for the purpose to be achieved; and iii) the evil inflicted must not be disproportionate to the evil avoided.[68]

---

[60] *Ibid* at 190.
[61] *Ibid* at 197 (family law perspective) and 203 (criminal law perspective).
[62] *Ibid* at 192. See also a similar discussion at 202-3.
[63] *Ibid* at 255.
[64] *Ibid*. See also his Lordship's summary at 258-9.
[65] *Ibid* at 219ff.
[66] *Ibid* at 240.
[67] *Ibid* at 239.
[68] *Ibid*.

As a statement on Justificatory Necessity in English law, *Re A (Children)* should be approached with caution. Each judge alluded to the applicability of the necessity principle to the situation,[69] but only one firmly grounded their legal conclusion on it as a defence, and all were at pains to point out that their decision was being made in the context of a unique set of circumstances. To that end, the reasoning of the court suggests that there was no intention to create a broader rule beyond the instant case.[70] Further, the fact that each judge took a different line of reasoning to reach the same conclusion has served to obfuscate any kind of broader principle that might be derived even if there had been such an intention. Indeed, despite all three judges agreeing that the operation was in the best interests of Jodie, only Walker LJ argued that the operation would also be in the best interests of Mary as it would grant both children the bodily integrity they had been denied at birth. Both Ward and Brooke LJJ disagreed that the operation would be in the best interests of Mary but felt that in the resulting clash of duties owed by doctors/balancing of evils[71] it should be Jodie's interests that prevailed.

Thus, the same facts were decided on the basis of three different interpretations of the necessity principle: necessity as private defence (Ward LJ); necessity simpliciter (Brooke LJ); and necessity as a 'best interests intervention' (Walker LJ). Brooke LJ's formulation of a three-part test certainly encourages the extrapolation of a general lesser evils defence, but neither Ward nor Walker LJJ seems to have endorsed Brooke LJ's judgment on the legal aspects of the case. Nevertheless, Stark has suggested that insofar as a general justificatory version of the necessity defence might be distilled from Brooke LJ's judgment, one could establish some key differences which separate it from duress of circumstances. Both defences require there to be an absence of alternative actions open to the defendant, that the threat must be aimed at the defendant or someone close to them, and that the accused should have acted with reasonable firmness and without voluntarily placing themselves into the situation. However, Stark points out that only duress of circumstances requires a threat of death or serious injury and a degree of immediacy of the threatened harm's occurrence based on Stephen's definition of Justificatory Necessity endorsed by Brooke LJ.[72]

---

[69] Specifically, Ward and Walker LJJ spoke of the killing of Mary as being the 'lesser evil': *Ibid* at 192 per Ward LJ and 255 per Walker LJ.

[70] Brooke LJ specifically noted that Sir James Stephen's concerns that people would too readily avail themselves of exceptions to the law was rather unfounded since this was "an exceptionally rare event": *ibid* at 240.

[71] The former view being held by Ward LJ, and the latter view being held by Brooke LJ.

[72] Stark, *op cit*. at 958-9. See also the comments by DC Ormerod in *R v Jones* [2005] Crim LR 122 at 125.

Stark suggests that these distinctions would have implications for the application of a Justificatory Necessity defence. First, the absence of a threatened harm requirement creates the possibility that, unlike duress of circumstances, Justificatory Necessity can apply to acts which prevent damage to property.[73] Second, the absence of a requirement for the threatened harm to be immediate suggests that actions need not be conducted under pressure to have a Justificatory Necessity defence. However, this second proposition about lack of immediacy appears to be redundant. Considering the requirement that the accused must have had no reasonable alternatives to committing the offence, its absence changes little. Stark suggests that the defendant's actions in Justificatory Necessity "need not be conducted under pressure", and that significant time for reflection should not, as in duress of circumstances cases, cause the accused to lose the defence.[74] However, it is difficult to envisage circumstances outside of medical necessity cases (of which a majority will be classified as best interests interventions anyway) where a defendant would have significant time to reflect on the situation, and yet no real alternative course of action open to them.

In any case, it is still unclear if the Justificatory Necessity variation of the defence even exists. Stark is ultimately dismissive of utilising Brooke LJ's judgment to determine the contours of Justificatory Necessity on the basis that *Re A (Children)* is better thought of as a case involving defence of others, based on a clash of duties.[75] Attempts had been made to apply the defence, and the reasoning of Brooke LJ in *Re A (Children)*, in the tragic case of Tony Nicklinson, a patient with 'locked-in syndrome' who applied to the court to allow doctors to help him die.[76] The application was rejected on the basis that the court did not think it was their place to attempt a balancing exercise of the respective 'evils' at work in voluntary euthanasia cases – an argument Stark refers to as the 'democracy problem'.[77] Thus, Stark argues that Justificatory Necessity remains "just as suspect a defence as it always has been".[78]

---

[73] Stark, *op cit*. at 959.
[74] *Ibid*.
[75] *Ibid* at 960. Leverick reaches a similar conclusion: *Killing in Self-Defence* (2006), Ch. 1 generally. Although cf. Ormerod in *R v Jones* [2005] Crim LR 122 at 125 who suggests that *Re A* establishes a Justificatory Necessity defence as a competent plea to a charge of murder. See also *R v Thacker* [2021] 1 Cr App R 21 at para 91 where Lord Brooke's definition was cited with approval (although the court then proceeded to conflate justificatory necessity with duress of circumstances in the following paragraphs).
[76] *R (on the application of Nicklinson) v Ministry of Justice* [2012] EWHC 2381 (Admin). There was a co-applicant, referred to as Martin, who also sought a similar declarator on assisted suicide, rather than voluntary euthanasia.
[77] Stark, *op cit*. at 961-3.
[78] Stark, *op cit*. at 961.

3.4 Immediate Threat/Danger of Death or Serious Injury

We now turn to Hume's requirements and their application to the necessity and coercion defences in Scots law. The remainder of this chapter will examine Hume's first and most important requirement; that the accused commit the crime to avoid an immediate danger of serious injury to themselves or to others.[79] There are several elements to unpack. First, a note about terminology. In *Thomson*, Lord Wheatley highlighted a distinction between the 'threats' and 'danger' in coercion cases, stating that both must be immediate.[80] It would seem that the term 'threats' here should be interpreted as shorthand for commands uttered by a human agent, with harm contingent on a failure to comply, as opposed to the more general sense to mean a person or thing likely to cause harm (i.e. as a synonym for danger). Thus, the coercer threatens the accused (threat) by saying they will seriously injure or kill them (danger) should they fail to comply. It could be argued that a requirement for both the threat *and* the danger to be immediate suffers from redundancy – a threat in this context is incomprehensible without reference to the danger to which it relates, and the current construction which separates these terms suggests that there may be cases where, although a danger of death or serious injury may be imminent, the plea would fail on the basis that the corresponding threat was either missing or came too early in time.[81]

3.4.1 Implied threats

Separation of 'threat' and 'danger' also has ramifications for implied threats: i.e. demands accompanied by non-verbalised threats which are inferred from the circumstances. The Scottish courts have never had to consider whether an implied threat may suffice, but the answer to this question is intrinsically linked to our definition of the term 'threat'. If, for example, Scotland chose to follow the Canadian approach in admitting implied threats,[82] it would become difficult to justify using the term 'threat' in anything

---

[79] *Lord Advocate's Reference (No. 1 of 2000)* 2001 JC 143 at 159, confirming that necessity could be pled in defence of others.

[80] *Thomson v HM Advocate* 1983 JC 69 at 77.

[81] Take for example the scenario where an accused does not take the initial threat seriously until the coercer initiates the violence which was previously threatened. Presumably we would say that the coercer's subsequent actions may constitute the threat, or alternatively that the threat subsists, but doing so constitutes an implied admission that the term 'threat' is synonymous with 'danger' and has no real relevance or meaning in coercion cases.

[82] *R v Mena* (1987) 34 CCC (3d) 304 at 320, applied in *R v McRae* (2005) 199 CCC (3d) 536. A similar approach has also been taken in Ireland: *Attorney General v Whelan* [1934] IR 518. In English law, the courts have considered that the indirect relaying of a threat to the accused may suffice (i.e. 'hearsay duress'):

other than the more general sense, as synonymous with danger. If all that is required is a danger of physical harm, absent an explicit 'do X or else' from a human threatener, the line between coercion and necessity becomes very blurry. The facts of the Canadian case of *R v McRae*[83] serve as a useful illustration of this point. The accused assisted his cousin to bury the corpses of two hitchhikers that the latter had killed and was convicted of being an accessory after the fact to murder.[84] On appeal, the court held that his defence of duress was successfully made out on the basis that, as they were in a remote location, the cousin had a gun, and he had executed one of the two hitchhikers in the accused's presence, McRae had strong reasons to believe that his cousin would harm him if he refused to comply despite the absence of an explicit threat.

Of course, one can immediately see the similarities this factual scenario shares with cases of necessity. In both situations the accused is placed in a dangerous situation where their options are limited to a choice between embracing that danger and the commission of a crime. Indeed, the only real distinction between cases like *McRae* and one such as *Tudhope v Grubb*[85] – where a man pled necessity to a charge of drink driving where he was attempting to flee three men seeking to seriously injure him – appears to be that McRae's cousin, the aggressor, requested the accused's help and he chose to comply. Perhaps there is something to be said about the psychology of a command in such contexts (and whether or not it is backed by a threat) which makes the distinction an important one in terms of the accused's guilt; or perhaps its importance lies in its ability to move the finger of blame from the accused to the threatener, an important ability given our intuitions about crime and justice often demand *someone* to blame for a perceived injustice. Indeed, the previous chapter demonstrated that a distinction based on the presence of the commands or design of another human agent has been recognised throughout history.[86]

Nevertheless, under the current rules of Scots law it would be incoherent to give McRae a coercion defence while still maintaining a distinction between threats and danger as in *Thomson*.[87] Further, given that both defences are said to share the same foundation

---

R v Brandford [2017] 4 WLR 17, although cf. commentary by K Laird: [2017] Crim LR 554 at 556. See also *R v N* [2007] EWCA Crim 3479.

[83] (2005) 199 CCC (3d) 536.

[84] Notably, the Scottish equivalent of art and part liability does not recognise liability for involvement in offences after the fact because the accused must be shown to act in concert with one another and it is clear that McRae had no common purpose to murder the hitchhikers: *McKinnon v HM Advocate* 2003 JC 29. A similar scenario taking place in Scotland would likely result in charges of attempting to pervert the course of justice: see, e.g. *Murphy v HM Advocate* 2013 JC 60.

[85] 1983 SCCR 350. For an equivalent in English law, see *R v Willer* (1986) 83 Cr App R 225.

[86] See section 2.4.3.

[87] Irrespective of whether the scenario could be classified as coercion or necessity, a person in McRae's position in Scotland may nevertheless have no defence on public policy grounds, as it is unlikely that either

and rules, it is contended that pleading the necessity defence would be more appropriate in cases where a threat is implied. If a person reasonably perceives that they are in danger, and provided that the other elements of the defence are made out, it would be pedantic and indeed unjust for the law to determine their guilt on the basis of whether particular words or phrases were spoken. Strict adherence to one particular type of defence, here coercion and the threat requirement, could lead to issues of nomination in terms of specificity – e.g. is it sufficient if the threatener demands an accused rob a bank, or must it be a nominated bank?[88] Such discussions are unhelpful and serve only to distract from the important elements of the defence, such as whether the accused reasonably believed they were in danger.

The biggest barrier to a successful plea of necessity in such cases is establishing a strong connection between the threatened harm and the offence charged, sufficient to prove a close factual nexus that establishes a constrained choice. Although as yet undiscussed in Scottish courts, this point came under scrutiny in the English case of *R v Cole*,[89] where the court had to consider the accused's claim of duress in answer to charges for robbing two building societies. Moneylenders had beaten the accused and threatened him and his family with further violence if he did not expediently repay the debt he owed them. Here there was no implied threat – the lenders had made it explicit what would happen if Cole failed to pay in a timely manner – but they made no suggestion about how Cole was to find the money to repay them. The trial judge considered that neither duress by threats nor duress of circumstances could be made out since there was no direct connection between the threats and the offence committed. In reaching this conclusion, he referred to the line of authority establishing duress of circumstances, stating that in each of those cases a direct link could be made between the threats and the offences committed to avert them.[90]

As a result of the trial judge's reasoning, Cole changed his plea to guilty and was sentenced to six years' imprisonment. He appealed and the question for the court was whether an unnecessarily strict view of duress had been taken. In assessing Cole's situation the appeal court quickly dismissed the possibility of duress by threats, rightly noting that for such a defence the threatener must nominate the specific crime(s) to be committed.[91]

---

defence can be successfully pled to charges of (being a party to) murder: see fn. 127 below. This does, of course, depend on precisely how such a case would be charged given that crimes after the fact are not recognised in Scots law.

[88] See A Reed, "Duress and Specificity of Threat: Commentary on *R v Z*" (2003) 67 *J. Crim. L.* 281 at 283.

[89] [1994] Crim LR 582.

[90] *Ibid*. On duress of circumstances generally, see above at section 3.3.1.

[91] *Ibid* at 583.

On the heading of duress of circumstances the appeal court also agreed with the trial judge, going further to state that in determining what nexus must exist between the threatened peril and the offence committed in order for the defence of duress to arise, particular regard must be had to the existence of an imminent peril. It was this imminent peril that cases such as *Willer*, *Conway* and *Martin* featured, and the court suggested that the offence committed by the accused as a result of the threat should be "virtually spontaneous" for duress of circumstances to be considered.[92] In a demonstration of just how fine a line there can be in these cases, the court questioned whether *Martin* should be regarded as an example of duress by threats rather than duress of circumstances.[93] They ultimately concluded, however, that despite the 'imperfect logic' inherent in the concept it did not matter which applied; what was important was that an imminent peril forced the actor to respond in a spontaneous way.[94] Despite admitting a degree of immediacy and peril in Cole's situation, the court nevertheless felt that it "fell short of the degree of directness and immediacy required".[95]

The conclusion is therefore that necessity may well operate to include cases of *prima facie* coercion with an implied threat that would otherwise fail, so long as the other requirements are strictly adhered to and some strong factual nexus can be established between the offence committed and the perceived threat by the accused. In practical terms, this strong factual nexus will be established if the conduct can be said to arise 'immediately' in response to the threat, understood as a spontaneous action. This aspect of the judgment in *Cole* may be regarded as unduly strict – it was accepted by the court that Cole had mere hours to find the money – and the suggestion appears to be that if Cole had immediately gone to rob the building society after being beaten he would have had a valid defence, which seems unlikely. Rather, what separates Cole's situation from the other successful cases of duress of circumstances was access to alternative courses of action. He could have gone to the police for assistance or moved his family to a temporary safe location until the danger had passed. It is for this reason, not the lack of imminent peril, that Cole's plea deserved to fail.

Thus, a distinction between threats and danger in coercion cases may be sustained while maintaining individual justice in cases which feature implied threats by pleading necessity. This is because, in contrast to coercion cases which emphasise the importance of

---

[92] *Ibid*.
[93] *Ibid*. They stressed that the case was, however, correctly decided.
[94] *Ibid*.
[95] *Ibid*.

the distinction between threats and danger, necessity cases are incapable of being understood other than in the context of threats in a general sense – whether they are human in origin or otherwise. There is therefore some scope for confusion when discussing the case law and its application to each defence. *Moss* states that *Thomson* and all other coercion cases apply to the law of necessity, but it is redundant to speak of the threat needing to be as immediate as the danger in such scenarios, as the court does in *Thomson* in relation to coercion. For the above reasons, it would therefore provide the most clarity to refer exclusively and universally to the *danger* as the subject of immediacy. In other words, in both necessity and coercion cases, the danger must be immediate to ground the plea.

3.4.2 Danger must be of death or serious injury

The danger an accused avoids must be one of death or serious injury, such that other emotional or financial harms will be insufficient.[96] Tadros has suggested that *Cochrane v HM Advocate*[97] leaves open the possibility of lesser threats sustaining a plea of coercion.[98] He bases this proposition on the fact that the court in *Cochrane* had pointed out that Hume left open the possibility of compulsion where the crime committed was not atrocious.[99] This logic, however, appears to operate on a conflation of the threatened harm with the offence committed – Hume leaves open the possibility of compulsion being available for lesser *offences*, he says nothing of lesser *threats*. In fact, a stronger argument might have been to point to the fact that the threats in *Cochrane* were to blow up the accused's home and physically assault him if he did not comply, and the court chose to dismiss both threats on the grounds that they were too remote, rather than on any notion that blowing up a house was an insufficient threat.[100]

There are, nevertheless, several reasons why it is unlikely that a Scottish court would uphold such an argument. First, while the court in *Cochrane* did not immediately dismiss the threats for lack of relevancy, it would be incorrect to say that they were therefore approving of them. It is clear that some scepticism existed regarding the authenticity of the threats, and indeed a discussion about the credibility of the threats made

---

[96] *Thomson v HM Advocate* 1983 JC 69 at 72 per Lord Hunter (approved by Court of Appeal); *Lord Advocate's Reference (No 1 of 2000)* 2001 JC 143 at 158, para [42].
[97] 2001 SCCR 655.
[98] V Tadros, "The Structure of Defences in Scots Criminal Law" (2003) 7 *Edinburgh Law Review* 60 at 67-8.
[99] *Ibid* at 68. See also Alison, p.672.
[100] 2001 SCCR 655 at 662.

up a large portion of the judgment.[101] It is thus unclear how much, if any, stock can be put into their failure to discard them for lack of relevancy. Second, the case of *Lord Advocate's Reference* has already established that the possibility for other types of harm to ground the necessity defence has been ruled out.[102] If, as has been outlined above, *Moss* is to be regarded as synchronising the rules for the necessity and coercion defences, then it would appear that the question has a foregone conclusion – only threats of death or serious injury will suffice. Finally, there is a clear desire by the Scottish courts to keep both defences within narrow bounds,[103] and thus there is a strong policy motivation to restrict the defences to a limited kind of threat.

Nevertheless, there does appear to be one exception that courts have been willing to entertain: that is threats of a sexual nature. In *D v Donnelly*,[104] the accused was charged with drunk driving and claimed she had only done so to escape (further) sexual assault.[105] Her necessity defence failed on the basis that there was no 'immediate threat of death or serious injury' at the time of her driving, but the court specified that "immediate danger of sexual assault of the type described here would qualify, if the circumstances merited it".[106] There have been no further cases on this point, and it is unclear whether the phrase 'of the type described here' in the judgment should be taken to imply that there are some types of sexual assault which would not warrant a defence. Without further clarification, it seems likely that any conduct deemed to be sexual in nature would suffice, provided it could be regarded as an immediate threat.[107]

3.4.3 The threat must be extraneous to the accused

Although no Scottish court has considered this question, note should be made to the English case of *R v Rodger and Rose*,[108] which concerned whether the threat must be extraneous to the accused, or otherwise derive from an external source. The two defendants had attempted to break out of prison after the duration of their sentences had been increased, and they claimed duress on the basis that they had become suicidal and would have killed themselves if forced to stay. The court held that, even if their accounts were

---

[101] This aspect of the case is discussed in detail in the next section.
[102] *Lord Advocate's Reference (No 1 of 2000)* 2001 SCCR 297 at 314 (2001 JC 143 at 158).
[103] *Moss v Howdle*, at 126.
[104] 2009 SLT 476.
[105] The accused was described in the case as having been 'indecently assaulted' throughout the judgment.
[106] 2009 SLT 476 at 477.
[107] This point is discussed in more detail below in chapter nine.
[108] [1998] 1 Cr App R 143.

accepted, no defence was available owing to the fact that the threats were not extraneous to the accused, and that to accept such threats would introduce an entirely subjective element divorced from any extraneous (i.e. objective) influence.[109] The court considered such an extension to be unacceptable as it would "amount to a licence to commit crime dependent on the personal characteristics and vulnerability of the offender".[110] In *Quayle*,[111] a case concerning the cultivation and use of cannabis to alleviate chronic pain, Lord Justice Mance confirmed the importance of an objective test to duress and necessity, stating that a requirement that the belief be 'well-founded' "imports a need for it to have been manifested externally and an ability to measure and assess it accordingly".[112]

3.4.4 A danger to whom?

The Scottish authorities on necessity have held that the threatened harm does not have to be directed towards the accused, with the court of appeal in *Lord Advocate's Reference* going as far to say that:

> "there is no acceptable basis for restricting rescue to the protection of persons already known to and having a relationship with the rescuer at the moment of response to the other's danger."[113]

Thus, provided the other requirements are made out, it seems that the defence is available even where the person who is threatened with harm is a stranger to the accused. The cases of *Docherty v HM Advocate* [114] and *HM Advocate v McCallum*[115] suggest that a coercion plea may be grounded in threats to the accused's family. Taking these cases together with the approach in *Lord Advocate's Reference* to necessity, it is therefore likely that threatened harm to a stranger would suffice for coercion cases too.[116] English law is somewhat consistent and stricter in its approach to both defences holding that duress may be pled where the threats are made against members of the accused's family,[117] and that

---

[109] *Ibid* at 145.
[110] *Ibid*.
[111] *R v Quayle* [2005] 2 Cr App R 34. Discussed in more detail below at section 9.2.
[112] *Ibid* at para. [47].
[113] *Lord Advocate's Reference (No 1 of 2000)* 2001 JC 143 at 159, para [44].
[114] (1976) SCCR Supp 146.
[115] (1977) SCCR Supp 169.
[116] Chalmers and Leverick suggest that, even under an excuse construction of the defence, it is "surely possible" that a person might be placed in the necessary state of fear for the defence to hold: *Defences* at para. 5.13.
[117] *R v Hasan* [2005] 2 AC 467 at 490, para. [21] per Lord Bingham; *R v Brandford* [2017] 4 WLR 17 at 23, para. [32].

duress of circumstances may be pled where the harm is threatened to a person for whom the accused "reasonably [regards] himself as being responsible".[118] The situation is less clear when neither the threat nor the harm are directed at the accused – e.g. where A threatens B with violence unless he commits act X and C, overhearing, commits act X on behalf of B – but such cases appear to be *prima facie* instances of necessity and thus, given the approach taken in *Lord Advocate's Reference*, valid (assuming the other requirements are made out).[119]

3.4.5 The offending act must be proportionate and capable of removing harm

In *Lord Advocate's Reference* the court made several comments about the necessary quality of the offending act committed by the accused for a defence, based on a close relationship between that act and the danger it averts. Lord Prosser stated that this should be interpreted as requiring the accused to have had reason to think that the acts carried out at the material time had some prospect of removing the threatened danger.[120] If the action would achieve no more than a postponement or interruption of the danger, the court thought that the necessity assessment could become quite difficult and involve issues of proportionality, as merely making a danger less likely "might not be regarded as justified by necessity at all".[121] Further, Lord Prosser went on to define proportionality in the following paragraph as the requirement that "the conduct carried out must be broadly proportional to the risk",[122] adding that this was always a question of fact to be determined in the circumstances of the particular case.

Generally speaking, the reasonable prospects test will be an easy one to meet, as in the majority of cases the relationship between the conduct and the threatened danger it averts will be obvious – the court in *Lord Advocate's Reference* gave the example of damaging a runaway vehicle to prevent it from causing greater damage.[123] One can think of countless other examples too, such as stealing a fire hydrant to put out a fire, or robbing a bank under instruction out of fear of death. Rather, what this test appears to be aimed at

---

[118] *R v Shayler* [2001] EWCA Crim 1977 at [63]. See also *R v Abdul-Hussain* [1999] Crim LR 570 at p.7 of the transcript, and the commentary in *R v Quayle* [2006] Crim LR 148 at 152.
[119] The strength of such a claim may well depend on whether one treats necessity as a justification or an excuse – while a case can be made under both constructions, the former certainly provides a stronger foundation.
[120] 2001 JC 143 at 159, para [46].
[121] *Ibid*.
[122] *Ibid*.
[123] *Ibid*.

is limiting the applicability of the necessity defence in acts of civil disobedience.[124] The facts in *Lord Advocate's Reference*, for example, were that the accused had damaged a ship related to the Trident nuclear missile programme. They claimed their actions were done out of necessity as the government was breaching international law by keeping such weapons to the danger of human life. Irrespective of whether the other requirements for a necessity defence were made out, it was unlikely that the damage caused would have any real impact on either the capabilities of the weapons, or the government's policy in this area. The appeal court therefore considered that, as the defence is concerned with finding a direct connection between the offending conduct and the avoidance of a particular danger, if there was no reasonable prospect of the offending conduct removing that danger then this necessary relationship between the two would not be established and there could be no defence.

However, it is unclear why this conclusion should be couched in the language of proportionality, particularly given that its use here refers to no more than an act's likelihood of removing a specified risk. Lord Prosser's comments appear to conflate the prospects of an action eliminating a risk (relevancy) with an assessment of comparative risk (proportionality). The terms are not synonymous. In reality, they refer to separate questions related to the quality of the offending conduct, and a proportionality assessment properly understood would likely fail in a majority of these civil disobedience style cases that the term is being employed. Indeed, the factual context of *Lord Advocate's Reference* itself demonstrates that such a comparison makes little sense because the damaging of a ship, i.e. property damage, pales in comparison to the potential harm from a nuclear war. The accused's actions were clearly proportionate to the threatened harm, but had little to no chance of removing that harm.

One might suggest that proportionality here refers to a comparison of two realities: the present one where the offence was committed, and a hypothetical one where it was not. In both realities the threat of nuclear annihilation is and will continue to be present, but in the present reality there is an additional harm of damage to property. Thus, the hypothetical reality in which the damage to property does not occur is to be preferred because there is 'less harm', and no defence can be claimed. Such a theory would make sense of Lord Prosser's claim that a lack of such proportionality would cause an accused's act not to be

[124] See, e.g. SM Bauer and PJ Eckerstrom, "The State Made Me Do It: the Applicability of the Necessity Defence to Civil Disobedience" (1987) 39 *Stanford Law Review* 1173 at 1178ff; F Stark, *op cit* at 961ff. (discussing the 'democracy problem').

"justified" by necessity at all.[125] Of course, this conclusion implies that necessity is a justificatory defence in Scots law, which appears to be at odds with how either defence has been understood both before and after this case.[126]

This kind of reasoning would, however, necessarily imply that the unlawful act must be *effective* at removing the harm for there to be a defence. Indeed, if the rationale is based on lesser evils reasoning, it becomes important that there is, in fact, 'less evil'. This would deny the defence to a person who made a reasonable error of judgment about the likelihood of their actions having an impact, and thus likely take the reasonable prospects requirement beyond what it was intended for: as a mechanism for limiting defences in cases where an accused seeks to expose or embarrass the government for perceived injustices. It is therefore submitted that the mention of proportionality in Lord Prosser's discussion about the reasonable prospect requirement should be treated as superfluous, and that this test should be considered separately from any proportionality requirement that applies to reactive defences.

Indeed, and notwithstanding any complex relationship between reasonable prospects and proportionality, the concept of a proportionality requirement itself poses problems for an understanding of necessity (or coercion) as an excuse, for similar reasons expressed above. The concept of proportionality is inherent in the context of lesser evils reasoning where there is a balancing of harms to determine whether the unlawful act was, all things considered, permissible. However, as each defence is restricted to threats of death or serious injury, and further since it is unclear whether either defence can be pled against charges of murder,[127] it would be inaccurate to suggest that the defences are based on this kind of value judgement, as the test would appear to be redundant in most cases. One can make the argument that the term 'serious injury' is vague enough to invite issues of proportionality – e.g. where A maims B under the threat of having his legs broken by C – but this heavily restricts the value of the proportionality test to an ethical question which would have been decided by the jury in any event.

---

[125] *Ibid*.
[126] See section 3.2.1 above.
[127] In *Collins v HM Advocate* 1991 JC 204, the trial judge made *obiter* comments to the jury that coercion was not a defence to murder, but no case has arisen directly on this point. In contrast, necessity has successfully been pled against a murder charge in the unreported case of *HM Advocate v Anderson* (2006), with Lord Carloway stating that necessity is a complete defence to murder, culpable homicide and assault (PR Ferguson & C McDiarmid, *Scots Criminal Law: A Critical Analysis* (2nd edn, 2014), at para. 21.4.6). There is, however, no authority from the appeal court on this issue.

The inclusion of such a test also has ramifications for (and indeed places limitations on) Hume's second requirement, that the violence must have been irresistible such that it dominated the mind of the accused.[128] This is because while proportionality asks us to consider whether the act was reasonable, an irresistible violence test seeks to determine whether an ordinary person of reasonable firmness would have succumbed to the threatened harm. There is thus the potential for a clash where the irresistible violence test points to a defence, and the proportionality test does not. This does not create an impossible choice, presumably the proportionality test would win out and no defence would be available, but it does so at the expense of a rational basis for the defence where it is unclear precisely *why* the accused is being held blameworthy. Clearly, the evidence points to normative foundations for the defences, but this cannot be on the basis that the law recognises leniency for an overborne will.

Despite the above concerns, the court has repeated this proportionality test on several other occasions, with Lord Wheatley stating in *Thomson* that "a fine balance may have to be struck between the nature of the danger threatened and the seriousness of the crime, calling for a value judgment",[129] and more recently in *Van Phan v HM Advocate*, where Lord Carloway endorsed Lord Wheatley's statement as correct.[130] In *Thomson* the test was raised in a discussion about whether coercion extended to murder cases, and likewise the court in *Van Phan* discussed it in relation to whether a human trafficking scenario might displace the standard strict immediacy requirement, so there appears to be a tendency to raise the issue of proportionality when the court is alluding to the potential for special circumstances which might displace the ordinary rules in favour of the accused, but it should be stressed that the only occasion on which the proportionality test has been in any way relevant involved limiting the scope of necessity to the accused's detriment.[131]

3.5 Conclusion

The first half of this chapter established a connection between the substantive defences of necessity and coercion in Scots law, setting the parameters for treating the case law on each defence as being mutually applicable. It also explored the English duress of circumstances defence, an important task given its influence on the Scottish understanding

[128] Explained in more detail below: see chapter four generally.
[129] 1983 JC 69 at 78.
[130] *Van Phan v HM Advocate* [2018] HCJAC 7 at [43].
[131] It should be noted that in the underlying case of *HM Advocate v Zelter* upon which the *Lord Advocate's Reference* was raised the accused were acquitted, but this was not on the basis of proportionality.

of necessity. After undergoing these clarifications, the latter half of this chapter introduced the most important of Hume's requirements as it has been received in modern Scots law – the unassailable requirement that the accused must have been acting to avoid an immediate danger of death or serious injury. Problems associated with the current understanding of these defences have been highlighted; namely how the characterisation and rationale of these defences creates difficult questions related to implied threats and issues of proportionality. Chapter four shall conclude Part I of this thesis by continuing to explore the remaining requirements applicable to the necessity and coercion defences in Scots law.

4. Contemporary Scots Law on Necessity and Coercion: Hume's Other Requirements

4.1 Introduction

This chapter shall continue where the last left off; by continuing the critical analysis of the remainder of Hume's requirements, as well as examining further tests which have developed in the modern era and their application to the defences of necessity and coercion in Scots law. We begin with a section on the immediacy requirement, a rule which forms part of Hume's first requirement that the accused be acting to avoid an "immediate danger of death or great bodily harm" but has become so complex that it deserves separate attention. This section shall explore the close connection found between the temporal requirement and Hume's second requirement for an inability to resist the violence, including a comparison of English law in this area which historically had a more lenient temporal requirement.

With this close relationship established, the next section shall go on to examine Hume's second requirement in more detail, focusing on what an inability to resist violence entails for both the quality of the accused's response, and the quality of the danger being avoided. This will lead into a discussion about subjective and objective approaches to the accused's assessment of their situation, including an analysis of the requirement that the threat must have dominated the mind of the accused, and any relevant characteristics which may affect this assessment.

Finally, this chapter shall complete its exposition of the modern law of necessity and coercion in Scotland by examining Hume's third and fourth requirements and how they have been interpreted into contemporary law. Particular attention shall be paid to the emergence of a new requirement that the accused must not be at fault for placing themselves into the dangerous situation, which includes a detailed look at voluntary association in coercion cases and how this might affect the availability of a defence in Scots law. Once the modern Scots law on these defences has been outlined, this thesis will have a suitable foundation from which it can explore more theoretical issues, such as the rationale for each defence, and the principles behind strict adherence to a temporal requirement.

4.2 The Immediacy Requirement

The last chapter sought to clarify precisely *what* must be immediate, but the immediacy requirement itself requires some exposition. Precisely what does it mean for the danger to be 'immediate'? The term immediate implies a degree of temporality; specifically, and in the context of a threat, it implies that the danger should be about to happen when the accused commits the offence. In practice, this translates to the proposition that the danger must refer to present and not of future injury.[1] In *Thomson*, for example, the appeal court rejected the appellant's claim that future injuries should suffice for coercion, with Lord Wheatley stating that only an 'immediate danger of violence' would suffice, as both the danger and the threat had to be immediate.[2] Likewise, in *Cochrane v HM Advocate*, threats to blow up the accused's house and beat him, made when the threatener and accused were at the residence of the victim, were held too remote to successfully ground the plea.[3] In *Lord Advocate's Reference*, Lord Prosser took this line of reasoning a step further in the necessity context by implying that an immediacy requirement required the actions of the accused to be undertaken spontaneously, as a result of urgent danger.[4] This suggests a connection between immediacy and the requirement that the accused's will be overborne – specifically that the former might act as evidence of the latter.

4.2.1 Immediacy and an inability to resist the violence

Returning to *Thomson*, Lord Wheatley suggested that 'immediate danger' would have to be construed in the circumstances, such that factors like the existence of reasonable alternative courses of action may suggest a lack of immediacy.[5] Chalmers and Leverick have therefore suggested that the immediacy requirement rules out a defence where the accused had a reasonable opportunity to avoid the danger and/or compliance by taking

---

[1] AM Anderson, *The Criminal Law of Scotland* (Bell & Bradfute, Edinburgh; 1892), p.6.
[2] 1983 JC 69 at 78.
[3] 2001 SCCR 655 at 661-62. There was also some scepticism over the immediacy of the threat of violence based on the accused's suggestion that even if he had fled the threatener would "probably have caught up later on": *ibid* at 662.
[4] *Lord Advocate's Reference (No 1 of 2000)* 2001 JC 143 at 178, para [100]: "What they did… was not a natural or instinctive or indeed any kind of reaction to some immediate perception of danger, or perception of immediate danger".
[5] *Thomson v HM Advocate* 1983 JC 69 at 77.

some other, lawful, course of action.[6] This construction is supported by the court in *Lord Advocate's Reference*, where Lord Prosser stated that,

> "It is clear that timing is a crucial consideration… Unless the danger is immediate, in the ordinary sense of that word, there will at least be time to take a non-criminal course, as an alternative to destructive action. A danger which is threatened at a future time, as opposed to immediately impending, might be avoided by informing… some responsible authority of the perceived need for intervention."[7]

This logic is, however, flawed because the reasoning followed in this line of cases is predicated on a conflation of two separate requirements: immediacy and an inability to resist the violence. As Chalmers and Leverick rightly point out, if the immediacy requirement exists to prevent an accused from claiming a defence where they ignored a reasonable alternative (and lawful) course of action, then it suffers from redundancy since the absence of a reasonable alternative course of action is already a separate requirement of the defence in Scots law.[8] We also know, from the reasoning of the cases above, that the immediacy requirement serves to limit the kinds of threats and conduct that will warrant a defence; this is a separate concern from whether the accused had other, reasonable options open to them to avoid the harm.

4.2.2 Is the immediacy requirement arbitrarily strict?

Thus, it is submitted here that the immediacy requirement is best understood as demanding that the threatened danger be temporally close to the wrongful act undertaken by the accused for a defence to be competent. This means that where the danger is no longer temporally close, the accused must desist from breaking the law or they will no longer have a defence.[9] This formulation of the immediacy requirement has, however, attracted criticism and has placed Scots law at odds with other jurisdictions.[10] In his

---

[6] Chalmers & Leverick, *Defences*, at paras 4.12 and 5.14 for necessity and coercion respectively.
[7] 2001 JC 143 at 157, para. [37].
[8] Chalmers & Leverick, *Defences*, at para. 4.12.
[9] *D v Donnelly* 2009 SLT 476; *Ruxton v Lang* 1998 SCCR 1; *McLeod v MacDougall* 1989 SLT 151. See also the English case of *DPP v Jones* [1990] RTR 33.
[10] The Canadian Supreme Court in *R v Ruzic* [2001] 1 SCR 687 held that the strict immediacy and presence of threatener requirements under s17 of the Canadian Criminal Code (RSC 1985, C-46) providing for a defence of compulsion by threats to principals of an offence infringed s7 of the Canadian Charter of Rights and Freedoms (Constitution Act 1982) in relation to life, liberty and security of person because they excluded threats of future harm to the accused or third parties. They therefore struck them down as being unconstitutional: see at paras [90] and [101]. English law is discussed below. Although cf. the US, specifically New York, where the requirement has been strictly construed: e.g. *People v Brown*, 333 NYS 2d

famous analogy looking at temporal requirements in American jurisprudence, Robinson argues that if a ship's crew notice a slow leak after leaving port, one which will not pose a danger for several days, and the captain refuses to return to shore, they are justified in committing mutiny because once the leak does start to threaten the integrity of the ship it will be too late to do anything about it.[11] It cannot be expected that the crew should just accept their fate. A focus on temporal considerations also raises concerns about the quality of the options available to the accused. This is because a focus on temporality tends to belie the factual context, such that in many cases protection by the authorities may be *possible*, but certainly not reasonable or effective in real terms.[12]

However, in the recent case of *Van Phan v HM Advocate*,[13] Lord Carloway suggested that the need for a "threat of immediate serious violence" was dependent on the context of a single violent act, to be distinguished from situations where the offence was drawn out over a longer period of time, and the circumstances meant that the normal recourse to the forces of law was unavailable.[14] Van Phan alleged that he was a victim of human trafficking who had been forced to cultivate cannabis by his captors out of fear of violence. Thus, in strict factual circumstances the defence may yet be available to those who have been kept prisoner while being forced to work in illegal industries such as drug operations (e.g. production and supply) and prostitution (e.g. controlling prostitution[15]), despite the absence of an immediate threat.[16] In the context of the above decisions, the judgment in *Van Phan* therefore appears to reiterate the importance of an immediacy requirement where the crime committed is a single act of violence. However, where the offence is not of this nature, immediacy of danger will still be important, but may be displaced where the standard reasonable alternatives (e.g. police protection) did not exist (because, for example, the accused was being held hostage). This decision represents a move away from the view that temporal requirements are important in and of themselves, towards one which values temporal requirements as part of a broader assessment of the opportunities the accused had to resist the violence.

---

342 (Sup. Ct. 1972) (where inmates held civilians and guards hostage in protest against deplorable conditions – the court rejected the necessity defence on the basis that the injuries feared were not imminent).

[11] PH Robinson, *Criminal Law Defenses* (West Publishing Co, 1984) (July 2021 Update), §124(f). These were broadly the facts of *United States v Ashton* 24 F. Cas. 873 (CCD Mass 1834) (No 14,470) which resulted in a successful plea of "justifiable self-defense against an undue exercise of power" (at 874, per Story J).

[12] See, e.g. Tadros, *op cit*. at 68; Robinson, *op cit*. §177(e). This issue will be returned to below in more detail when considering the requirement that the accused could not resist the violence.

[13] [2018] HCJAC 7.

[14] *Ibid* at 18.

[15] See, e.g. *R v LM, MB, DG, BT and YT* [2010] EWCA Crim 2327, in particular at para. [24]ff.

[16] Although Van Phan's plea was unsuccessful. This case is discussed in more detail at section 8.5.

4.2.3 The English approach to temporality

The approach taken in English law does, however, seem to have moved in the opposite direction, making it worthy of note. Initially, the Court of Appeal in *R v Hudson*[17] placed emphasis on the ability of the accused to resist the violence by rejecting the claim that the danger itself must be immediate. The case concerned two young women who were due to testify in court as witnesses but, prior to the court date, they were threatened by a man to keep quiet and thus decided to comply to avoid any harm. This decision was strengthened when they took the stand to give evidence and saw the man who had threatened them watching from the public gallery. Their perjury was subsequently discovered and they were charged. The court held that it was necessary to determine whether the threat was 'present and immediate' in the accused's mind at the time of the offence, sufficient to overcome the will of the accused such that the act was no longer voluntary.[18] If that threshold was met, the court felt that it did not matter if the danger of death or serious injury did "not follow instantly, but after an interval".[19] If this danger became too remote, however, the court thought it would be extremely difficult to prove that the threat was present and pressing on the accused's mind to the extent that their will had been overcome. In this sense, a temporal connection would be relevant insofar as it could provide credibility to the accused's account that their mind had been so dominated, with proof of such domination establishing that the accused genuinely believed they had no alternative than to commit the offence.

The approach to immediacy of threat taken in *Hudson* was followed for many years.[20] The reasoning of the case was applied with approval by the House of Lords three years later in *DPP for Northern Ireland v Lynch*,[21] and again in *R v Lewis*,[22] but in 1994 the court in *R v Cole*, as discussed above,[23] emphasised the need for an 'imminent peril' to ground the plea of duress of circumstances.[24] The reader will recall that the court in *Cole* considered both duress by threats and of circumstances and considered that, in line with the trilogy of cases establishing duress of circumstance and culminating with *Martin*,[25] there

[17] [1971] 2 QB 202.
[18] Per Lord Widgery at 206.
[19] *Ibid* at 207.
[20] The court's approach to the assessment of the reasonableness of alternative courses of action was also followed in a line of cases: *R v Baker and Ward* (1999) 2 Cr App R 335, *R v Lyness* [2002] EWCA 1759, and *R v True* [2003] EWCA Crim 2255.
[21] [1975] AC 653.
[22] (1993) 96 Cr App R 412 at 415.
[23] [1994] Crim LR 582, discussed at section 3.4.1.
[24] Although the court never challenged *Hudson* on this point.
[25] (1989) 88 Cr App R 343, discussed at section 3.3.1.

must be a direct and immediate connection between the threatened harm and the offence committed.[26] Lord Justice Simon Brown, who also delivered the judgment in *Martin*, stressed the need for the conduct to be "virtually a spontaneous reaction" to the threats.[27] In this sense, the kinds of threats that would put the accused in such 'imminent peril' would be severely restricted.

Nevertheless, and five years later in *R v Abdul-Hussain*,[28] the Court of Appeal disagreed with the observations made in *Cole* insofar as that ruling had neglected to draw a sufficient distinction between 'immediacy' and 'imminence' and was inconsistent with *Hudson*, with which the present court not only felt bound to follow but also preferred.[29] The appellants had hijacked a plane as part of an attempt to flee the Iraq regime that would have killed them. In language similar to *Hudson*, the court accepted that while an 'imminent peril' should operate on the mind of the accused so as to overcome their will at the time of the offence, the execution of the threat need not be immediately in prospect.[30] Lord Justice Rose put it thus:

> "[A]lthough the judge was right to look for a close nexus between the threat and the criminal act, he interpreted the law too strictly in seeking a virtually spontaneous reaction. He should have asked himself, in accordance with *Martin*, whether there was evidence of such fear operating on the minds of the defendants at the time of the hijacking as to impel them to act as they did and whether, if so, there was evidence that the danger they feared objectively existed and that hijacking was a reasonable and proportionate response to it."[31]

Finally, and despite the comments made in *Abdul-Hussain*, in the 2005 decision of *R v Hasan*[32] the judgment in *Hudson* was criticised by Lord Bingham on the basis that it weakened the requirement that the danger must be reasonably believed to be imminent and immediate.[33] Thus, an objective approach to the assessment of immediate danger was favoured. In addition, his Lordship cited Glanville Williams' comment that the decision in *Hudson* was "indulgent", and said he was sceptical that a witness in court would have no

---

[26] [1994] Crim LR 582 at 583.
[27] *Ibid*.
[28] [1999] Crim LR 570.
[29] These statements are not mentioned in the Criminal Law Review report but are available in the official transcript.
[30] [1999] Crim LR 570 at 571: "If Anne Frank had stolen a car to escape from Amsterdam and been charged with theft, the tenets of English law would not, in our judgment, have denied her a defence of duress of circumstances, on the ground that she should have waited for the Gestapo's knock on the door."
[31] *Ibid*, from official transcript.
[32] [2005] 2 AC 467.
[33] *Ibid* at 494.

opportunity to avoid complying with a threat of violence.[34] In contrast to the conclusions of the court in *Abdul-Hussain*, which he curiously described as being bound by the decision in *Hudson* but made no mention of the fact they were in *agreement* with it, Lord Bingham agreed with the decision in *Cole* and stressed the need for a degree of directness and immediacy in the context of the link between the danger and the offence charged.

This case appears to have represented a notable shift in the judicial consensus on the immediacy requirement.[35] Indeed, shortly after this decision Lord Bingham's judgment was followed in *R v Quayle*,[36] where the court held that the defence of duress of circumstances should be unavailable to persons charged with supplying cannabis for medicinal use on the basis that there was an insufficient degree of immediacy present, "recognised as a key element of duress defences" in *Hasan*.[37] The court admitted that *Abdul-Hussain* had established that the harm threatened need not be immediate but should be imminent, but said this statement had to be read in light of Lord Bingham's speech in *Hasan*.[38] This speech has since been cited and reaffirmed on various occasions.[39]

It is, however, not altogether clear where this precisely leaves the immediacy requirement in English law. For example, in *DPP v Mullally*[40] the court discussed the need to take evasive action such as seeking out and relying on the protection of the police, "if that option is realistically available".[41] This seems to fall into the same trap as Scots law, conflating an immediate danger with an ability to evade that danger. Likewise, in *R v N* the court thought that the decision in *Abdul-Hussain* (including, presumably, the facts of the case) was in line with Lord Bingham's decision in *Hasan*, and that *Hudson* should be regarded as a decision decided on its facts.[42] These cases therefore suggest that the 'reasonable belief' in the 'imminent peril' required may yet be interpreted generously to the accused, but considerations like the effectiveness of police protection and other evasive factors may play a role in deciding whether the case displayed the necessary 'imminent peril'.

---

[34] *Ibid*, citing G Williams, *Textbook of Criminal Law*, 2nd edn (1983) at p.636.
[35] Although see *R v Shayler* [2001] 1 WLR 2206 which implied a stricter focus on immediacy as a fundamental ingredient of the necessity defence.
[36] [2005] 2 Cr App R 34.
[37] *Ibid* at 152.
[38] This particular point comes from another report of the case: [2005] 1 WLR 3642 at 3670, para. [46].
[39] *R v Hussain* [2008] EWCA Crim 1117, paras. [17]-[18]; *R v Batchelor* [2013] EWCA Crim 2638, paras. [12]-[14]; *R v Brandford* [2017] 4 WLR 17 at 23, para. [33].
[40] [2006] EWCA Crim 3448 (Admin).
[41] *Ibid* at para. [8].
[42] [2007] EWCA 3479 at para. [12].

4.2.4 Summary: a close link between the immediacy and irresistible violence requirements

What should have become apparent from comparing the Scottish and English approaches to an immediacy requirement is the fact that Scots law conflates, rather than subordinates, the requirement with the separate rule that there be an inability to resist the violence. We saw above that in both *Thomson* and *Lord Advocate's Reference* the courts defined the immediacy requirement as ruling out a defence when the accused had a reasonable opportunity to avoid the danger and/or compliance by taking some other, lawful, course of action. This moves the requirement away from purely temporal concerns and suggests that its function is to assess the necessity of the accused's situation. This is a regrettable result as it makes understanding the separate requirement that the accused must have been unable to resist the violence difficult to comprehend.

In contrast, the English line of cases beginning with *Hudson* are, logically speaking, cases where the temporal requirement was held to be subordinate to the requirement that there be an inability to resist the violence threatened. In *Hudson* specifically the threatened harm to the accused was not immediate – they had no guarantees that, had they testified in court against their coercer's wishes, they would have suffered violence that night after leaving court. They may well have suffered the retaliatory violence at a later date. On a strict application of any temporal requirement then, their plea should have failed. That their plea did not fail demonstrates an acknowledgement of the fact that the defendants genuinely believed they had no alternative courses of action open to them (i.e. that police protection would be ineffective).

Nevertheless, it should also have become clear that the English approach is not without its own faults, and the case law here again demonstrates how difficult it can be to effectively separate temporal considerations from the assessment of an actor's potential opportunities for reasonable and legal evasion. In particular, it seems that discussions centred around the beliefs of an agent's ability to evade the threat invite confusion about whether immediacy is to be assessed as an objective fact or subjective concern. English law appears to take the view that a reasonable person would consider there was a threat, and at least some degree of temporal connection existed, but otherwise the assessment of an immediate danger is undertaken in the context of whether it can be established that the accused had an opportunity to evade the harm. Scots law, in contrast, appears to take an objective approach to all questions – the threat must have objectively existed, there must be harm which is temporally close to the threat (i.e. an immediate danger), and the accused must not have had any alternative course of action open, irrespective of how unrealistic

that option was. We shall now consider this latter requirement in more detail, in an attempt to untangle it from temporal considerations.

4.3 An Inability to Resist the Violence

In the previous section the immediacy requirement was examined in detail, and it became clear that it has historically been considered in close relation to the requirement that the accused was unable to resist the violence.[43] As a result, the rule on immediate danger has tended to be understood by Scottish courts as requiring more than just a close temporal connection, often requiring the threat to have overcome the accused's will and to demonstrate a lack of reasonable alternative courses of action. It is submitted that these considerations are better thought of separately, as factors which influence the assessment of whether the accused was able to resist the threat of violence. This section shall unpack precisely what this requirement entails in the necessity and coercion defences in Scots law, with every care taken to avoid the tendency to conflate this subject with the related, and yet distinct, temporal requirement.

Being concerned with mental turmoil sufficient to overcome a person's will, it is easy to think that the presence of an inability to resist the violence should be regarded as causing coercion and necessity to negate the *mens rea* of an offender. Indeed, this was the position taken by Lord McCluskey in *HM Advocate v Raiker*,[44] a case concerning coercion where, in his charge to the jury, he claimed that a real, genuine fear would negate the imposition of any "evil intention" such that the accused would lack "the criminal state of mind that is a necessary ingredient of any crime".[45] While there might be some credibility to this line of reasoning in relation to murder and the decision in *Drury v HM Advocate*[46] concerning the 'wicked' intention required, it is unclear that the same logic would apply for other offences where having *mens rea* simply refers to having basic intent, knowledge or recklessness towards a particular outcome, irrespective of motive.[47] The court in *Lord*

---

[43] The nomenclature used to identify this requirement is fluid, and thus commentators may refer to it in terms of whether an accused was able to evade the threat/violence/harm, or conversely in negative terms such that the rule is understood as being satisfied when the accused had no (reasonable) alternative course of action. Theoretically speaking, it is an elucidation of the necessity principle – that the accused must have had no option other than to do as they did. I do not subscribe to any particular formulation of the requirement, but for ease of reference I will refer to it as an 'inability to resist the violence' (and minor variations thereof, context depending) when discussing the rule in current Scots law.

[44] 1989 SCCR 149.

[45] *Ibid* at 154.

[46] 2001 SLT 1013.

[47] A point highlighted by Chalmers and Leverick: *Defences*, para. 5.04. Cf. *R v Hibbert* [1995] 2 SCR 973 at para [45] which appears to leave the door open for duress negating *mens rea* in Canadian law.

Advocate's Reference[48] came to a similar conclusion in relation to necessity, holding that whether or not the accused intended to inflict damage was not at issue: it was whether they were justified in so doing.[49]

Simply put, that the accused must have been unable to resist the violence to have a defence is an elucidation of the necessity principle that these defences embody.[50] To that end, it is better to understand these defences as involving *intentional* conduct. Generally speaking, the law does not endorse the retaliation and vengeance of citizens – protection of property and personal integrity being a function of the law – but it does allow persons to protect their own interests when the law is unable to fulfil this duty, and this concession is recognised through reactive defences like self-defence, necessity and coercion. Thus, as per *Thomson*, "where there is opportunity to seek and obtain the shield of the law in a well-regulated society, then recourse should be made to it", otherwise no defence will be open.[51] Likewise, Lord Prosser stated in *Lord Advocate's Reference* that the necessity defence was a "dispensing power exercised by the judges where they are brought to feel that obedience to the law would have endangered some higher value".[52] As we established above in the previous section, that higher value is the protection of bodily integrity from death or serious injury.

4.3.1 The absolute effectiveness of police protection

Retreat is a common example of an alternative course of action which provides the accused with an ability to resist the harm. If a person can retreat to safety from impending danger, they will not be entitled to rely on a defence if they decide to nevertheless carry out a criminal offence.[53] One variation of retreat which is very common and expected of persons in these situations is seeking out police protection. This point was directly at issue in *Trotter v HM Advocate*,[54] where the accused was charged with possession of drugs with

---

[48] 2001 JC 143 at para [31].
[49] See also *McNab v Guild* 1989 SCCR 138 at 143 where Gordon, in his commentary, points out that since recklessness in reckless driving depends on objective circumstances, circumstances of necessity must be capable of being relevant in the same way as any other circumstances. This, he says, incidentally shows that "necessity is a question of fact, of the *actus reus*, and not of *mens rea*". In English law, see *Santos v CPS Appeals Unit* [2013] EWHC 550 (Admin) which held that the defence of necessity could be pled against charges for strict liability offences.
[50] Discussed in more detail at chapter five below.
[51] 1983 JC 69 at 77. See, e.g. *McLeod v MacDougall* 1989 SLT 151.
[52] 2001 JC 143 at 155, para [33]. This kind of lesser evils reasoning is not without its own difficulties, as discussed in more detail below at 4.5 and chapter five generally, particularly at 5.2.5.
[53] E.g. *Dolan v McLeod* 1999 JC 32.
[54] 2000 SCCR 968.

intent to supply after heroin was discovered on his person while visiting his father in prison. He claimed that a third party had told him to take the drugs to his father and that if he refused, his father would be hurt. He had the opportunity to inform the authorities about this plan before visiting the prison but chose not to do so out of fear for his father's safety, wagering that the authorities would be unable to protect his father from retaliation.

The issue appealed was procedural in nature, but the appeal court commented that, irrespective of procedure, the defence was not made out on the facts because there was no immediate danger, proven by the fact that Trotter had the opportunity to inform the authorities of his predicament.[55] The court recognised that Trotter had not taken this opportunity because he was afraid, implying that he did not believe the police would be able to adequately protect his father from being harmed in prison, but nevertheless rejected his plea stating that such facts may only be taken into account in mitigation of sentence.[56] This line of reasoning was followed more recently in *Sinclair v HM Advocate*,[57] where Lord Brodie seemed to suggest that where the police were not contacted, coercion could only operate as a plea in mitigation.[58] The Scottish courts therefore can be seen to take a strictly objective approach to the question of whether an accused could have resisted the harm – provided that police protection (or other forms of retreat) merely exist, it will be assumed that such options were an effective, and therefore reasonable, alternative and it is unlikely that the requirement will be satisfied. This is a very high threshold as, in most cases, the mere possibility of contacting the police will always be an option.

This approach corresponds to the current approach taken in English law. The reader will recall from the last section that, historically, the court in *Hudson* was sympathetic to the fear experienced by coerced persons, and its potential to overcome their will such that it did not matter whether the threatened harm was immediate if it could be carried out at a proximate time. In relation to police protection, the court had rejected the Crown's claim that the accused should have made efforts to alert the police about the threats, pointing out that this would eliminate any distinction between cases where police protection would be effective and those where it would not.[59] This would, the court noted, restrict the defence to situations where the accused was kept in the custody of those making the threats (i.e. hostage scenarios), or where the time between issuing the threat and the commission of the

[55] *Ibid* at 972.
[56] *Ibid*.
[57] [2017] HCJAC 88.
[58] *Ibid* at para. [11].
[59] [1971] 2 QB 202 at 207.

offence made recourse to the police practically impossible.[60] The court felt that, despite wishing to keep the defence within narrow bounds, this would be too restrictive. Thus, while it would always be open to the Crown to prove that the accused had failed to avail themselves of some course of action that would have nullified the threat, the question had to be assessed based on the accused's genuine belief in the context of their age and circumstances, and to any potential risks that the alternative course of action might pose.[61]

However, we saw that *Hasan* has since rejected this reasoning, endorsing the 'practical impossibility' position that the defendant must reasonably believe that the harm threatened will follow (almost) immediately on their failure to comply with the threat, otherwise there will be "little if any room for doubt" that they could have taken evasive action, such as going to the police, and the defence will be denied.[62] In reaching this conclusion, Lord Bingham was at pains to highlight that any assessment of the age and circumstances of the defendant should not be used to determine questions about the efficacy of any evasive action open to them, seeing these as two separate questions to be determined by the jury.[63] Of course, this conclusion will in many cases render an assessment of the defendant's age and circumstances redundant if the plea is to ultimately fail on the basis of a subsequent (strict) objective assessment of the evasive actions open to them.

While one must keep in mind the public policy considerations influencing this position in both jurisdictions, it nevertheless seems unduly callous. Despite these defences (particularly coercion) being referred to as 'concessions to human frailty', the court essentially expects the average person to always blindly trust the police to protect their interests, even when social/environmental circumstances might give them good reason to distrust them or cast doubt on their ability to help.

4.3.2 The subjective threat versus the objective danger

Another interesting aspect to consider is how Scots law distinguishes between threats and danger for the purposes of the temporal requirement. Historically, this seems to have accounted for the distinction between the English contextual approach to the

---

[60] *Ibid*.
[61] *Ibid*. Despite the weakening of the *Hudson* decision generally in light of *Hasan*, the court's approach to the assessment of the reasonableness of alternative courses of action has been followed in a line of cases: *supra* fn.20.
[62] [2005] 2 AC 467 at 494, para. [28], per Lord Bingham.
[63] *Ibid* at para. [24].

assessment of alternative options taken in *Hudson*, where an emphasis was placed on the psychological impact of the threat on the accused, as opposed to the authenticity of the corresponding danger in Scots law. The trial judge in *Thomson*, for example, recognised that the threat must be such "that would overcome the resolution of an ordinarily constituted person of the same age and sex as the accused",[64] which is much in line with the reasoning of the court in *Hudson*. However, Lord Wheatley added on appeal that, while in both jurisdictions the threat must have dominated the mind of the accused, English law places the emphasis of determining the existence of such a domination on the immediacy of the threat, whereas Scots law places it on the immediacy of the danger arising out of that threat.[65]

Drawing attention to these distinct temporal stages implies that the focus in Scots law when determining whether the commission of a law-breaking act was inescapable is not on the subjective state of the accused (i.e. whether the actor in their circumstances thought they were unable to resist the threats). Rather, it suggests that Scots law understands such questions objectively, by limiting the assessment to consider only the contemporaneous actions that could have been taken to avoid the threatened harm. Thus, Lord Wheatley doubted that a Scottish court would return the same decision in *Hudson*, given the fact that "the basic question is whether there was immediate danger of the threat being implemented in the event of non-compliance *at the point in time when the decision had to be made*".[66] Pace Lord Wheatley then, if the danger itself does not require immediate action there can be no defence, because then the actor can take other legal routes to resolve their issue. Of course, this kind of language automatically precludes questions about the efficacy of police protection, because the definition of the immediacy requirement presupposes the answer.[67]

As a result of this reasoning, Scottish courts are unlikely to engage in subjective discussions about the reasonableness of some alternative courses of action. We already saw that the court in *Trotter* was unwilling to indulge the accused's fear that the police would

---

[64] 1983 JC 69 at 72. See also Anderson *op cit*. p.6; *Lord Advocate's Reference (No 1 of 2000)* 2001 JC 143 at 158, para [42].
[65] *Ibid* at 79-80. But see now *Hasan* [2005] 2 AC 467 for the contemporary English position: *supra* fn. 61.
[66] *Ibid* at 80 (my emphasis added). See also *Cochrane v HM Advocate* 2001 SCCR 655 at 662 where Lord Rodger discusses the authenticity of the accused's responses and available options to the threatened harm: "It is perhaps because, on the appellant's narrative, the threats appear to lack the kind of immediacy which Hume envisages that the question of the appellant's reaction to the threats comes into prominence".
[67] See, for example, the language used in *SB v HM Advocate* 2015 JC 289 at 297: "However, [necessity] can only arise where the parent acts in the face of an immediate danger of death or serious injury to the child... If there is no immediate danger, the parent would be bound to adopt a course of action which is not otherwise criminal".

be unable to protect his father had he informed them of the drug plot, and no mention was made to Trotter's circumstances, gender, age or otherwise. Likewise, in *Cochrane v HM Advocate*,[68] Lord Rodger highlighted that requiring the threat to have overcome the resolution of an ordinarily constituted person of the same age and sex as the accused was superfluous where there was also a requirement that the accused should have been in immediate danger of death or serious injury because any reasonable person would succumb to such threats:

> "In other words, in a situation where there was an immediate danger of death or great bodily harm, unless it could be averted the danger would ipso facto be regarded as being such as would overcome the resolution of any ordinarily constituted person, whatever their age or sex."[69]

His Lordship therefore saw the strict harm requirement as necessarily precluding much of the discussion about what kinds of danger might overcome the will of the accused and, since almost everyone can be assumed to rationally fear death or serious injury, discussions about an accused's age or gender etc. would be redundant. In other words, when inquiring as to whether the threats dominated the mind of the accused, the court is not evaluating the *types* of threats which might ground the defence, but rather assessing the *quality* of a limited type of threat as part of a broader determination of whether the threatened danger was truly irresistible.

4.3.3 Domination of mind as a separate requirement

Scots law's strict adherence to the temporal requirement has had the unfortunate side effect of creating an artificial distinction between assessments of the accused's inability to resist the violence on the one hand, and the associated assessment that the threatened harm dominate the mind of the accused on the other, leading to illogical results. In the case of *Dawson v Dickson*,[70] for example, a firefighter's plea of necessity to drunk driving charges after he drove a fire truck to clear the way for an ambulance carrying a person in critical condition was rejected on the basis that there was evidence that he would have driven the truck anyway, irrespective of whether an emergency required him to do so. Dawson had been drinking while off-duty with an on-duty colleague when the latter was

[68] 2001 SCCR 655.
[69] *Ibid* at 661.
[70] 1999 JC 315.

summoned to a serious road traffic accident, and Dawson decided to attend. On arrival at the station, two fire trucks were dispatched and the accused drove one of them to the scene of the emergency, despite other firefighters being present and qualified to drive the vehicles. After arriving at the scene of the accident, it became clear that one of the fire trucks was blocking the path of an ambulance carrying a patient in critical condition – Dawson was at that point the only firefighter in the vicinity who could operate the vehicle and so undertook to clear the obstruction, thereby causing a collision which prompted the criminal charges. Dawson was found guilty of drunk driving at the scene of the accident, with the court stating that Dawson's initial decision to drive to the scene demonstrated that circumstances of necessity did not inform his reasons for action and could not therefore be relied upon in his defence.[71]

This reasoning is problematic because it forced the court into a discussion as to whether there was a 'conscious dilemma', and whether it could be said that the accused contemplated a difficult choice.[72] The court decided that there was no such dilemma precisely because he would have driven the vehicle irrespective of the circumstances. This conclusion seems unconvincing – the circumstances in the qualifying moment were that Dawson had to move the truck as, if he did not, a casualty might die.[73] That is true irrespective of Dawson's prior motivations and it seems disingenuous to conclude, as the court did, that Dawson was somehow not acting based on the obvious dilemma facing him. There was both an immediate danger, as well as no other reasonable alternative course of action and thus, in that moment, Dawson's mind was dominated by the need to move the fire truck. If the courts were reluctant to allow Dawson to evade liability on public policy grounds because his prior conduct was reckless (which is entirely reasonable), it would have been more logically coherent to deny the defence on prior fault grounds as the accused had voluntarily put himself in a situation where he might need to rely on a necessity defence.[74]

Indeed, in assessing whether an 'ordinary person of reasonable firmness' would have been influenced to respond to the threats as the accused did, the law is asking whether the threats are credible and genuine.[75] If a small child demands a fully grown man should

---

[71] *Ibid* at 317. Cf. the English case of *DPP v Bell* [1992] Crim LR 176.
[72] *Ibid* at 318.
[73] The court noted that the driving to the scene was not libelled in the charge – only Dawson's driving at the locus of the accident was at issue: *ibid* at 316.
[74] E.g. *McNab v Guild* 1989 JC 72. This requirement is discussed below in more detail. In an ideal world, Dawson would also have been charged for his initial drunk driving to the scene, allowing the court to disallow the defence on the basis that when the offence originally occurred no immediate danger was present.
[75] See, e.g. *Lord Advocate's Reference (No 1 of 2000)* 2001 JC 143 at 158, para [42].

rob a bank or face violence, we expect the man to resist the threat because it has very little credibility. In *Dawson*, however, the threat of death to the casualty if the fire truck was not moved was very credible and would have compelled anyone in the accused's situation to react as he did, irrespective of the fact that he put himself in that situation in the first place by unnecessarily drink driving (which is a separate question). This aspect was discussed in some detail in *Cochrane*, where Lord Rodger highlighted a passage from Alison stating that a defence may be available "provided he did not yield too easily to intimidation, but held out as long as in such circumstances can be expected from a man of ordinary resolution".[76] His Lordship went on to highlight with approval the judgment of Murnaghan J in the Northern Irish case of *Attorney-General v Whelan*,[77] where the latter argued that threats of death or serious injury which were sufficient to overbear the ordinary power of human resistance "should be accepted as a justification for acts which would otherwise be criminal".[78]

4.3.4 The relevant characteristics of the person of 'reasonable firmness'

Indeed, the decision in *Cochrane* determines precisely which characteristics should be taken into consideration when deciding the credibility of a threat and the accused's ability to resist it. Lord Rodger was at pains to stress the importance of an objective test as a way of ensuring that people who are responsible for their actions cannot use the defence to avoid the consequences of those acts, simply because of some failing in their personality which they should be striving to master.[79] His Lordship considered that most other characteristics should only be of relevance when determining the matter of punishment.[80] Nevertheless, he recognised that the standards expected of younger persons would be different to those of older persons, and likewise for women as compared to men.[81] Chalmers and Leverick have suggested that *Cochrane* leaves open the question of what characteristics are relevant to this assessment, citing the following passage by Lord Rodger:

---

[76] 2001 SCCR 655 at 663, citing A Alison, *Principles of the Criminal Law of Scotland* (1832), p.673.
[77] [1934] IR 518.
[78] *Ibid* at 526.
[79] 2001 SCCR 655 at 666.
[80] *Ibid* at 669 and 671.
[81] *Ibid* at 667. This particular example appears to be made on the basis of an inequity of strength and thus it is perhaps more accurate to suggest that what is relevant is the accused's physical ability relative to the threatener, irrespective of gender.

> "… consider whether an ordinary sober person of reasonable firmness, *sharing the characteristics of the accused*, would have responded as the accused did. Therefore, in a case where the accused lacks reasonable firmness, the jury must disregard that particular characteristic *but have regard to his other characteristics*."[82]

However, it seems clear from Lord Rodger's judgment that precisely the age and sex (perhaps better understood as the physical ability) of the accused are the only relevant considerations (i.e. the 'other characteristics') for the judge and jury to consider in such cases.

Chalmers and Leverick have suggested that it remains possible for recognised psychiatric conditions to provide relevant characteristics on the basis of the above interpretation of *Cochrane*, but also on the basis that other cases, such as *Lord Advocate's Reference*,[83] leave open the possibility by vaguely referring to 'relevant characteristics' that the reasonable person 'shares with the actor'. They point to the judgment of *Bone v HM Advocate*[84] as providing further evidence of this proposition, on the basis that the court defined 'personal characteristics' in the context of whether a mother should have been able to protect her daughter from the acts of a violent partner as "the appellant's physical, social and psychological circumstances".[85] One issue with this logic, recognised by Chalmers and Leverick, concerns the fact that *Bone* defined personal characteristics within the context of determining whether an element of the *offence* had been made out, which can be contrasted with a definition of personal characteristics which is to be used in the context of reactive *defences*.[86] There have, however, been no further cases to date which deal with this issue.

4.3.5 Summary: the overbearing influence of the temporal requirement

Similar to the cases discussed above, *Cochrane* can also be accused of complicating and conflating the requirements in these assessments. Lord Rodger correctly stated that in order for the defence of coercion to be left to the jury, the trial judge must have been satisfied that the appellant's will had been overcome and that he had committed the offences charged as a result of threats constituting an immediate danger of death or serious bodily injury.[87] However, the subsequent treatment of this statement was to suggest

---

[82] *Ibid* at 670-1 (emphasis added), discussed in Chalmers & Leverick, *Defences*, para 5.22.
[83] 2001 JC 143 at 158, para [42].
[84] 2006 SLT 164.
[85] *Ibid* at 167.
[86] Chalmers & Leverick, *Defences*, para 4.19.
[87] 2001 SCCR 655 at 662.

that Cochrane's will could not have been overcome precisely *because* there was no immediate threat.[88] This is the logical conclusion of a framework which asks the question of whether the accused "did genuinely believe that [the threatener] would severely assault him and blow up his house *then and there*".[89] The ever-present temporal requirement automatically dismisses as irrelevant any threats which are incapable of being carried out in the instant moment.

The temporal requirement can therefore be said to: i) limit the types of threats that can be considered relevant; as well as ii) reduce the scope of the irresistible harm requirement such that it is subordinate to the point of redundancy. This strict temporal requirement allowed Lord Rodger to reach the conclusion that, as with Scots law on coercion, the English law on duress includes "an objective test, referring to the response of a sober person of reasonable firmness, sharing the characteristics of the accused",[90] despite the clear difference in how this test was applied in each jurisdiction at the time. English law was not constrained by a strict temporal requirement, allowing the court in *Hudson* to apply the modified objective test precisely as it should be – rather than ignore the test completely where the circumstances do not conform with a strict temporal limitation. To be clear, insofar as the circumstances *do* conform with there being an immediacy of danger, e.g. the archetypal example of a person who is forced to commit an offence at gunpoint, *Cochrane* makes clear that the Scottish courts will undertake an objective assessment to determine whether the ordinary person of reasonable firmness who shares the same age, gender and physical ability as the accused would have succumbed to the threats, provided also that those threats are of death or serious injury, but this is the extent to which Scots law allows any kind of subjective discussion on the matter.

4.4 Hume's Third and Fourth Requirements

In *Thomson*, Lord Wheatley briefly discussed Hume's third and fourth requirements and their application to modern Scots law. With regard to both having a

---

[88] *Ibid*: "One of the threats was to blow up the appellant's house and… Cannon could not have done that immediately since… Cannon was in the complainer's house at the time. The other was a less specific threat to hammer or kick the accused… Even here, however, the appellant said that he thought that, if he had run away, Cannon would probably have caught up later on. This suggests that the appellant was really envisaging an attack in the future. It is perhaps because… the threats appear to lack the kind of immediacy which Hume envisages that the question of the appellant's reaction to the threats comes into prominence. Be that as it may, we must deal with the appeal on the basis that the danger embodied in the threats could properly be regarded as being sufficiently immediate to give rise to the defence."

[89] *Ibid* (my emphasis added).

[90] *Ibid* at 664.

backward and inferior part in the perpetration of the offence, as well as disclosure of the fact and restitution of the spoil on the first safe and convenient occasion, he considered these to be unable to affirm or disprove that the accused was acting under coercion in and of themselves, being better thought of as factual elements which would go to assessing the credibility and reliability of the accused's account that they were coerced.[91] Thus, in modern cases considerations which touch on these statements have sometimes become relevant to discussions about whether there was a threat of death or serious injury or death at the material time which the accused could not resist,[92] but they are not requirements in themselves.

4.4.1 Other issues relating to the credibility of the accused

One further issue that emerges in relation to credibility is whether any prior fault on the part of the accused affects their ability to plead a defence. Specifically, if the accused has done something prior to the unlawful act – such as voluntarily joining the gang who coerced them to commit the charged offence, or placing themselves in a dangerous environment that thereafter necessitated breaking the law to escape – does the accused's voluntary exposure to the risk of harm negate their ability to claim that they acted under duress or from necessity? In Scots law there has been just one case that considered the issue of prior fault directly.[93] In *McNab v Guild*[94] the appellant became involved in an altercation in a car park with two men, forcing him to flee in his car onto the main road and collide with another vehicle. He was charged with reckless driving and pled necessity. At trial his plea was rejected on the basis that he had admitted to previously leaving the car park when the initial threats were made only to return shortly afterwards, and thus had voluntarily placed himself in the dangerous situation and could not rely on that danger to remove his guilt.

---

[91] 1983 JC 69 at 78. Lord Carloway endorses this statement in *Van Phan v HM Advocate* [2018] HCJAC 7 at [42]. The requirement that the accused play a backwards and inferior role has limited application to necessity cases anyway.

[92] The fourth requirement in particular, specifically the question of whether the accused desisted from the criminal activity at the first opportunity, has been of relevance in a number of necessity cases where the accused drove while over the legal alcohol limit to escape a danger: see *supra* fn. 9.

[93] I argued above that *Dawson v Dickson* 1999 JC 315 should have been considered on this issue: see section 4.3.3.

[94] 1989 JC 72.

On appeal the court held that for the appellant to have a defence there needed to be an immediate danger of death or serious injury to the appellant at the material time.[95] This was apparently lacking on the facts of the case. However, the appeal court also suggested that the bare fact that McNab had returned to the scene did not, in and of itself, provide grounds to reject a necessity plea. They noted that while his return to the scene "certainly suggest[ed] that he had driven back in order to confront his co-accused again", in the absence of further proof the sheriff was wrong to conclude that the defence of necessity was foreclosed.[96] The implication is therefore that the court *would* have rejected a necessity defence if it could have been proved that McNab had intended to re-enter the dangerous circumstances, as opposed to returning some time later hoping that his assailants had left. This *obiter dictum* should not, however, be taken as an end to the matter; it remains a live issue.

Other than *McNab v Guild*, the only other case which discusses this point in any capacity is *Thomson*, where the trial judge Lord Hunter suggested that:

> "if the accused has joined an association where such threats from associates and the dangers arising from them are reasonably to be expected, a defence of coercion by such associates will not avail him."[97]

This point was not of relevance to the appeal, and thus received no commentary from Lord Wheatley.

Chalmers and Leverick suggest that, whether one conceives of necessity and coercion as justifications and/or excuses, the rationale of each defence would seem to support a prior fault rule. Under both defences, where an accused seeks to excuse their conduct (i.e. wrongful conduct for which the accused deserves no *blame*), there is logic in restricting each defence to situations where the accused is free from fault in creating the situation.[98] In terms of conduct which might be justified, there are public policy concerns which appear to necessitate a prior fault rule in coercion cases, for fear that the defence would be abused in cases of organised crime.[99] As a justification, Chalmers and Leverick note that because necessity involves the infliction of harm to an innocent third party's

---

[95] The court applied the reasoning of the English duress of circumstances decisions to reach this conclusion: *ibid* at 75.
[96] *Ibid* at 75-6. It appears that the presiding sheriff had made unfounded assumptions about how long an interval there was before McNab returned to the car park, as well as McNab's reasons for returning.
[97] 1983 JC 69 at 73.
[98] Chalmers & Leverick, *Defences*, paras 4.20 and 5.23.
[99] *Ibid* at 4.20. Note that this argument applies to both justificatory and excuse versions of the coercion defence.

interests, it may be difficult to rationalise this harm as justified if the accused was somehow at fault for creating the dilemma in the first place.[100] The reasons which support a prior fault rule where necessity and coercion are considered justifications are certainly weaker – it is not difficult to conceive of circumstances where, despite bringing about the dilemma, commission of a crime is a lesser evil[101] – but the Scottish courts have never rationalised either defence in justificatory terms.[102]

4.4.2 Prior fault and voluntary association in other jurisdictions

In the absence of jurisprudence on this issue in Scots law,[103] analogies might be drawn from English and Canadian law. In *Perka v The Queen*[104] the issue was discussed directly in relation to necessity. The appellants were charged with importing cannabis into Canada after their ship, destined for Alaska and carrying the plant as cargo, was forced to seek refuge on Canadian shores when several mechanical failures and poor weather conditions struck. They pled necessity in answer to their charges, arguing that they had been forced to bring the drug into Canada against their will. Dickson J, delivering the opinion of the Supreme Court and outlining the Canadian law of necessity, began by classifying the defence as an excuse.[105] On this basis, necessity could be understood as a defence where the accused's actions were normatively 'involuntary'.[106] Thus, if the necessitous situation was clearly foreseeable to a reasonable observer, or if the actor contemplated or ought to have contemplated that their actions would likely give rise to an emergency requiring breaking the law (irrespective of the underlying legality of those

[100] *Ibid* at 4.21.
[101] Chalmers and Leverick question the authenticity of the public policy concerns raised in relation to coercion being utilised as a 'criminal's charter', on the basis that such threats would, in many cases, not be genuine: *Defences*, para 5.23.
[102] See sections 3.2.1 and 3.3.2 above, outlining necessity in Scots law as a transplant of the English duress of circumstances, which in turn is understood in excusatory terms. See also section 4.5 below.
[103] The only case that seems to come close to examining this topic is a sheriff court decision concerning the 'special reasons' defence to disqualification under s34 of the Road Traffic Offenders Act 1988: *Lowe v Mulligan* 1991 SCCR 551. In this case the accused had driven to a pub, parked their car on a double yellow line, and then later moved the car to a nearby car park after consuming alcohol. He was charged with driving with excess alcohol in his blood, but claimed there were special reasons for not disqualifying on the basis that not only was the car illegally parked, but the next morning the street would become one-way to accommodate roadworks and thus the car would become a hazard. This argument was accepted by the trial judge, who paid little attention to the Crown argument that the accused knew about the potential obstruction before he parked. It is also unclear why the accused did not park in the car park to begin with, given that it was only a short distance away. Cf. *Mackay v MacPhail* 1989 SCCR 622. It is unlikely that this logic would stand to scrutiny in a necessity case where the remedy sought was full exoneration rather than a reduction of the penalty.
[104] [1984] 2 SCR 232.
[105] *Ibid* at 246-8.
[106] *Ibid* at 250-1. Discussed in more detail at section 5.2.4.

actions), then the response could not be said to be 'involuntary' in the relevant sense, and would preclude a defence.[107]

English law has discussed the issue in some detail in relation to duress, a discussion which has revealed several ways of understanding a prior fault requirement. Initially, in *R v Baker and Ward* Roch LJ took the position that the defence should be denied only when it was reasonably foreseeable that the defendant would be placed under pressure to commit the type of crime they committed.[108] Thus, and on this view, where a person joins a gang that sells drugs, they would have a defence if they were then pressured to commit armed robbery.[109] In *R v Heath*,[110] by contrast, Kennedy LJ relied on an earlier Court of Appeal judgment[111] to hold that where the defendant voluntarily, and with knowledge of its nature, joined a criminal organisation which they knew might bring pressure on them to commit an offence, they could not rely on the duress defence. It was, according to the court, "the awareness of the risk of compulsion which matters".[112] *R v Harmer*[113] took this a step further, with Goldring J interpreting *Heath* to mean that duress would be ruled out where the defendant voluntarily exposed themselves to "unlawful violence", irrespective of their membership to a criminal group or that they did not foresee being asked to commit a crime.[114] Here the defendant owed money for drugs, but had no other affiliation to his coercers.

Finally, in *Hasan*[115] Lord Bingham considered the line of authority dealing with this issue and concluded that if a defendant voluntarily surrendered their will to the domination of another, there could be no defence.[116] Thus, the focus of a prior fault test was on the relationship between the defendant and their coercer, as per *Heath*. He added that "nothing should turn on foresight of the manner in which… the dominant party chooses to exploit the defendant's subservience".[117] Lord Bingham therefore concluded that *Baker and Ward* was wrongly decided. In assessing the level of foresight required by the defendant, he said that an objective test was appropriate – they ought reasonably to

---

[107] *Ibid* at 256.
[108] (1999) 2 Cr App R 335 at 344 and again at 346.
[109] Chalmers & Leverick, *Defences*, para. 5.23.
[110] [2000] Crim LR 109.
[111] *R v Sharp* (1987) 85 Cr App R 207.
[112] [2000] Crim LR 109. This statement is found in the official transcript and cited in *R v Hasan* [2005] 2 AC 467 at 498 per Lord Bingham.
[113] [2001] EWCA Crim 2930.
[114] *Ibid*. See, in particular, paras 9-13 and 16 of the judgment.
[115] [2005] 2 AC 467.
[116] *Ibid* at 498, para [37], and 499, para [39].
[117] *Ibid* at para [37].

know that they may be the subject of compulsion.[118] The law here now appears to be settled, with Lord Bingham's judgment being followed recently by the Court of Appeal in *R v Brandford*,[119] the court adding that acting "out of love, infatuation or under pressure" could not invalidate a person's choice to voluntarily associate with those engaged in criminal activity.[120]

4.4.3 The importance of voluntary association to prior fault

From the above analysis several observations can be made about the (potential) Scottish approach to questions of prior fault and voluntary association. Applying the English jurisprudence to Lord Hunter's statement in *Thomson* gives the impression that Scots law would take a similar approach to the question of voluntary association as did Lord Bingham in *Hasan*. Lord Hunter identifies the accused's voluntary association with a criminal group as being a pre-requisite, as well as implying that an objective test would determine whether the accused should have realised the criminal nature of that association, on the basis that such conduct was to be "reasonably expected".[121] Chalmers and Leverick suggest that, as per the line of English authority which holds that the defendant's association must be truly voluntary,[122] Scots law would likely hold a defence open to an accused whose association was coerced. Likewise, they have suggested that, in the absence of Scottish authority on the point, Scots law would presumably follow English law in allowing persons who had ceased association with their coercers prior to being compelled to commit an unlawful act a defence.[123]

However, and with respect to other examples of prior fault, the Scottish courts have proven to be stricter in their application of the necessity defence. *McNab v Guild* has suggested that where it can be shown that an accused has knowingly placed themselves in (or created a) dangerous situation, it is unlikely that a defence will be open to them. Likewise, while the appeal court in *Dawson v Dickson*[124] did not introduce the concept of prior fault to their decision, the facts there were analogous as Dawson had voluntarily (and

---

[118] *Ibid* at 499, para [38].
[119] [2017] 4 WLR 17.
[120] *Ibid* at para [45].
[121] *Thomson v HM Advocate* 1983 JC 69 at 73.
[122] Chalmers & Leverick, *Defences*, para 5.25, citing *R v Sharp* [1987] 1 QB 853 at 861, per Lord Lane CJ and *R v Harmer* [2001] EWCA Crim 2930 at [16].
[123] *Ibid*, citing *R v Baker and Ward* [1999] 2 Cr App R 335 at 346 and *R v Lewis* (1993) 96 Cr App R 412 at 417.
[124] 1999 JC 315. See above at section 4.4.1.

quite unnecessarily) placed himself into circumstances where he would be forced to break the law due to the urgency of the situation. The court there did deny Dawson a necessity defence, but they came to this conclusion on the basis that his prior actions had demonstrated that the necessity of the situation had not dominated his mind, and thus there was no dilemma. Nevertheless, the substance of this decision is analogous to the Court of Appeal's decision in *Harmer* in suggesting that the accused must not voluntarily expose themselves to a threat of violence generally. It is therefore unclear how a future Scottish court would attempt to reconcile these two lines of reasoning: *Thomson* suggests that Scots law would follow the English approach to voluntary association which is more favourable to the accused (requiring active association), but the few Scottish cases which deal either directly (*McNab v Guild*) or factually (*Dawson v Dickson*) with prior fault suggest a more stringent test which appears to be more in line with the reasoning in the earlier English authority *Harmer*.

One may be tempted to suggest that the reasoning in *Dawson* is not inconsistent with the understanding of voluntary association arrived at in *Hasan* – in both cases the court stressed that if a person had voluntarily made a prior choice which would make their subjection to a subsequent threatened harm foreseeable, a defence would not be open to them. Nevertheless, the conclusion in *Dawson* is clearly more consistent with the legal ruling in *Harmer*. To some extent the Scottish approach to a case of voluntary association may depend on how far the courts wish to unify the rules on coercion and necessity, with the reasoning in *Harmer* being more consistent with a view of these defences as embodying "a distinction without a relevant difference".[125] However, given the clear lack of judicial appetite for any wholesale analysis of these defences, this conclusion seems unlikely. It is much more likely that a Scottish court would opt to follow the established line of English authority, despite any discrepancies this may create between the treatments of voluntary association and prior fault.

4.5 Rationale for the defences

Finally, something brief should be said about the current understanding of the rationale for the defences of necessity and coercion in Scots law. As aforementioned,[126] the law currently differentiates between affirmative defences which assert the blamelessness of

[125] *R v Howe* [1987] AC 417 at 429, per Lord Hailsham.
[126] See thesis introduction at fn.9 and accompanying text.

the actor (excuses), or contest the wrongness of the act (justifications). Issues arise in relation to the necessity and coercion defences because, theoretically speaking, each is capable of being characterised as a justification or an excuse, depending on the context.[127] If A threatens to injure B unless B injures C, we might say that while injuring C was wrong, B's actions were nevertheless understandable given the circumstances, and B is therefore not blameworthy. In contrast, if B commits property damage to prevent a fire spreading which would have resulted in greater harm to both property and life, we might instead say that B's actions were permissible, or even positively correct: B is blameless, and therefore not an appropriate subject of criminal liability.

The Scottish criminal courts have taken an overly restrictive view of each defence, which has influenced their characterisation: as noted above the requirement for a threat of death or serious injury severely limits the circumstances in which a person can plead a defence,[128] such that B in the second scenario above would only have a defence of necessity if they could prove that the encroaching fire threatened life – greater harm to property would be insufficient. In general, however, it is not entirely clear how the Scottish courts characterise each defence. It seems likely that both are regarded as excuses, owing to the language used in the courts. For example, the trial judge Lord Hunter stated in *Thomson* that the "will and resolution of the accused must in fact have been overborne and overcome".[129] Likewise, the court in *Moss* stated that the act must have dominated the mind of the accused, such that if there is a reasonable legal alternative to disobeying the law, "then the decision to disobey becomes a voluntary one".[130] While far from conclusive, it would make little sense to hold that persons should be overcome with fear if the ground for exculpation is based on the conduct being correct. Equally, and as discussed above,[131] Lord Rodger has previously held that the Scottish law of necessity should be regarded as equivalent to the English law of duress of circumstances. The latter is understood as an excusatory defence in English law,[132] thus implying the same rationale for its Scottish counterpart. However, in the absence of any definitive statements from the courts, presumptions are the most accurate observations to be made.

---

[127] For academic support see, e.g., Chalmers & Leverick, *Defences*, at para 4.03, fn.15 and 5.04, fn.29 and accompanying text.
[128] See section 3.4 above.
[129] 1983 JC 69 at 72. See also *Docherty v HM Advocate* 1976 SCCR Supp. 146 at 146; *Sayers v HM Advocate* 1981 SCCR 312 at 319.
[130] 1997 JC 123 at 129. See also *Tudhope v Grubb* 1983 SCCR 350 at 352.
[131] See section 3.2.1. This point was later affirmed in *Lord Advocate's Reference* at para. [42].
[132] F Stark, "Necessity and *Nicklinson*" [2013] Crim LR 949 at 952-954.

4.6 Conclusion

Part I of this thesis has outlined the contemporary law of necessity and coercion in Scotland, examining the historical context and its evolution into modern law through the application of Hume's foundational requirements by the appeal court in *Thomson*. This chapter has continued and concluded this task, critically analysing the remaining requirements while focusing on the second major aspect of Hume's passage: the interlocking need for an overborne will, a lack of alternative options, and a strict temporal requirement between the offence and the threatened harm. It has been argued that the primacy of the latter requirement, that of temporal proximity, has acted as a bulwark to restrict the defence to only those circumstances where a threatened harm could be said to be immediate, to be contrasted from imminent or otherwise less temporally close harms. However, this strict adherence to an immediate danger has left the law in this area confused and unsatisfactory – it makes questions about alternative courses of action and aspects of the accused and their circumstances largely redundant, which is regrettable for defences which aim to redress the often harsh, objective nature of criminal offences. Indeed, we have examined jurisprudence from English law in *Hudson* which recognised these concerns, and helps to frame the Scottish reasoning in terms of distinguishing a psychological threat from an objective danger. It is regrettable that English law has followed in the footsteps of Scotland on this topic.

In the next part of this thesis, I will delve further into the rationale for these defences, examining their underlying nature and why we might wish to exculpate persons who break the law when confronted with situations of extreme pressure or individual emergency. It is hoped that by analysing these underlying philosophical concepts, a richer understanding of these kinds of defences can be adopted and potentially applied to Scots law.

PART II

5. The Normative Foundations of 'Reactive Defences'

5.1 Introduction

Part I of this thesis outlined the substantive defences of necessity and coercion as they are currently understood in Scottish criminal law doctrine as a way of introducing the topic and the relevant issues. Part II, beginning with this chapter, explores necessity and coercion from a theoretical perspective to discover what normative claims can be made about these kinds of defences and their role in a system of criminal law. The goal here is to establish an underlying normative basis for necessity and coercion to accurately assess the validity and effectiveness of the current rules and structure which apply to them. To this end, we must determine what situations these substantive defences are aimed at addressing, and on what basis these kinds of situations warrant a negation of criminal censure. Both necessity and coercion have traditionally been understood as exculpatory in nature, meaning that the defences serve to prevent the agent's responsibility for a *prima facie* wrongful act from transitioning to criminal liability, although it has not always been clear whether this is on the basis that the actor's conduct was all things considered 'right' or 'permissible', or rather that while the conduct was regrettable the actor was not blameworthy.[1]

The overall aim of this chapter is to determine exactly what, normatively speaking, is being recognised as affecting an actor's blameworthiness in a general, extra-legal sense when we say that they were 'coerced' or 'acted out of necessity'. To this end, two unique terms – 'situation of individual emergency' and 'situation of extreme pressure' – are developed here to collectively refer to those instances where a person commits an act ordinarily deserving of blame, owing to external pressures which play a decisive role in influencing their (ultimately) chosen conduct. This chapter will explore the precise definitions of, and distinction between, these two terms through an analysis of culpability theory to develop an alternative framework for understanding the coercion and necessity defences in Scots law and the distinction between them. By engaging in this normative

[1] RA Duff, *Answering for Crime* (Hart Publishing, 2009) at p.263ff.; see section 4.5. On the use of 'permissible' (as opposed to 'right' or 'good') see, in particular, p.266 (and below at section 5.2.1); see also S Uniacke, *Permissible Killing* (Cambridge University Press, 1994) at p.14.

analysis, it is hoped to tease out precisely what elements of these circumstances result in our intuition that a person who commits an offence in a pressurised situation should be treated more favourably than a person who commits that same act under normal circumstances.

As normative concepts, situations of individual emergency and extreme pressure are not bound by legal rules or principles. This necessarily entails that they are together broad enough to encompass not only circumstances which might warrant a necessity or coercion defence, but also other situations which encompass defences like self-defence. Indeed, both these normative concepts and the defences to which they apply are based on the concept of a reaction – a person reacting to the world around them. We can therefore develop a general category in the legal environment for defences which are based on this normative understanding of pressure and/or emotional turmoil. I denominate this category 'reactive defences.'[2] However, this thesis is solely concerned with how the concepts of situations of individual emergency and situations of extreme pressure can inform our understanding of the necessity and coercion defences – it does not consider their interaction with other reactive defences.

After establishing the distinction between situations of individual emergency and situations of extreme pressure and locating the latter concept in the realms of wrongful conduct, this chapter goes on to examine further what is meant by a 'constrained will' under extreme pressure, and how choice and capacity have traditionally been understood to affect an actor's culpability in this context. Here, the chapter explores alternative normative conceptions of situations of extreme pressure, resulting in a hybrid understanding of the concept which recognises both choice and character theories as being important considerations to determining culpability under a constrained will. In particular, a move away from treating either theory as exclusively explaining culpability allows us to better understand the role of emotions in this exculpation assessment, which is the subject of the following chapters. It is argued that a hybrid approach to culpability theories, perhaps counter-intuitively, allows us to present a unified theory of the role of emotions in situations of extreme pressure, which allows us to reach our normative conclusion in chapter seven.

[2] This term is inspired by the Scottish Law Commission who, in their discussion of a proposed overview of the necessity, coercion, self-defence and provocation defences, described such defences as being "reactive in nature": *Seventh Programme of Law Reform* (Scot Law Com No 198 (2005)), at para. 2.47.

5.2 Establishing a Normative Foundation for Reactive Defences

The goal of this section is to establish a normative foundation which can reasonably explain why we feel empathy for or approval towards those who commit criminal offences in pressurised situations. By doing so, we can more fully understand the basis of those defences in law which aim at negating liability because of emotional and/or circumstantial pressure(s) that the accused experienced at the time of the offence's commission, such as necessity or coercion. To be clear, I do not claim that the arguments being made here are revolutionary in altering how we think about pressurised situations and their handling by the law. Indeed, much of what follows should appear intuitive. Rather, my aim is more modest – it is to (re)emphasise these pre-established underlying normative truths about the ethical considerations of acting in pressurised situations, which can then be emphasised in the legal setting as a general baseline. From there, we can develop a framework that rationalises the requirements of defences like necessity and coercion.[3]

5.2.1 The nature of situations of individual emergency and extreme pressure

To begin, we must outline what it means for an actor to be faced with a 'situation of individual emergency'[4] or 'situation of extreme pressure'. Both are normative concepts which conceive of external states of affairs confronting an actor against their desires,[5] and which threaten one of their interests. They are predicated on the notion that, in order for a responsible act to take place, a person's body and mind must not be subject to external or internal pressure not initiated by the agent: their action must be unencumbered.[6] To this end we can add that if a person is aware of the wrongness of their action but cannot avoid it despite this recognition, it does not serve the purposes of the criminal law to hold that person responsible.[7]

---

[3] Although by its nature this framework will not be limited to these defences.
[4] This term is inspired by François Tanguay-Renaud, who uses the term 'individual emergencies' to cover situations where the criminal law fails to prevent an attack and individuals face great risks of harm for which they must react urgently if that harm is to be averted or minimised: F Tanguay-Renaud, "Individual Emergencies and the Rule of Criminal Law" in F Tanguay-Renaud & J Stribopoulos (eds), *Rethinking Criminal Law Theory: New Canadian Perspectives* (Hart Publishing, 2012), pp.21-54. Similar verbiage can also be found in Australian Commonwealth Criminal Code 1995, s10.3(1): "sudden or extraordinary emergency".
[5] The use of 'desires' here is deliberate as the more intuitive term 'will' has strong associations with mental fortitude, and I wish to solely focus on external circumstances here.
[6] W Hirstein, KL Sifferd & TK Fagan, *Responsible Brains: Neuroscience, Law, and Human Culpability* (MIT Press, 2018) at p.69.
[7] *Ibid* at p.178.

Traditionally, in the legal context these interests have been confined to physical integrity (i.e. a threat of death or serious injury), but for our normative purposes this could include other interests such as property. Thus, a person who damages a neighbouring house to create a firebreak and protect their own property would qualify. Equally, a scenario where a hiker is caught in a severe snowstorm and breaks into a nearby cabin to take refuge is *prima facie* both a situation of individual emergency and a situation of extreme pressure – the hiker is placed into the situation against their wishes, and one of their interests (here their physical integrity) is at risk.

The division of such scenarios into two separate categories is intended to recognise two distinct types of actor, as well as a corresponding difference in the normative attitude taken towards each. Situations of extreme pressure is likely to cover the majority of these types of cases – it envisages a person who, because of extreme pressure, is overcome by emotion, having some impact on their will, and acts on that basis. Normatively speaking we are unlikely to approve of the conduct but will see the actions as understandable. In contrast, situations of individual emergency cover those cases where a person acts to bring about what they believe is the least harm or a positive goal (i.e. emotions are not the driving force). Our focus here is on the external circumstances giving rise to their claim. Normatively speaking we may or may not approve of the conduct but, in any case, we view the actions as understandable.

To this end, situations of extreme pressure can be said to generally track excusatory defences, in that they rely on a concession to human frailty which views the actor's conduct as understandable for exculpation.[8] In contrast, situations of individual emergency may refer to those cases where the action is, all things considered, permissible and thus the existence of pressure, while still contextually important, does not hold the same exculpatory significance.[9] This concept therefore generally tracks justificatory defences. However, while these concepts may generally track excusatory and justificatory defences respectively, they are not intended to be exact matches. This is most notable in the case of situations of individual emergency where a person may act to protect an interest, but the action is deemed wrongful. Indeed, in the context of a normative framework situations of individual emergency only seek to explain why conduct undertaken within those parameters is *understandable*.

---

[8] On justifications and excuses generally, see the introduction at fn.9 and accompanying text.

[9] In this sense we might say, as Paley does, that fear is to be distinguished from 'necessity', in that it does not affect the unlawfulness of an act but may affect the wilfulness of an individual's conduct or the mental state of an actor: JM Paley, "Compulsion: Fear and the Doctrine of Necessity" 1971 *Acta Juridica* 205 at 207.

Such conduct is understandable because, at the very least, anyone else would have done as the actor did. It may be that anyone else would have done as the actor did precisely because it was objectively the right thing to do, but equally it might have been understandable despite its harmful nature, because it engaged with moral notions of what it is to be human which negate blameworthiness.[10] I choose to rely on the normative concepts developed here rather than the justification/excuse distinction primarily because it avoids the legal baggage associated with the latter, pre-existing terminology, but also in recognition of the difficulties associated with categorising defences like necessity and coercion as just one 'type' of defence. A bespoke distinction therefore seems appropriate, allowing us to examine reactive defences from their foundational starting point as a reaction, rather than in a purely act- or agent-centric manner as encouraged by the traditional justification/excuse distinction.

5.2.2 Differentiating intentional actions

Before evaluating the merits of this approach, we must first determine why we would ever treat two persons who commit the same offence differently when one of them acts within the context of the type of situation outlined above. We must therefore consider the kinds of circumstances which result in an actor committing an offence intentionally and yet being regarded as having done nothing wrong or, alternatively, as being undeserving of blame. In the context of a person who intentionally harms to avoid a negative consequence to their interests, there appear to be several ways of characterising this kind of conduct. It has been suggested, for example, that the voluntary nature of the actor's conduct in such scenarios has been impaired, in the sense that the actor lacks the ability to conform their behaviour to the law.[11] Thus, Robinson suggests that although an actor engages in such conduct voluntarily, and correctly perceives the nature of their act and that it is wrong, they may still be exculpated because they lack the capacity to control their conduct.[12]

Robinson comes to this conclusion because he regards exculpatory conditions as falling into one of only two categories – exculpation based on the harm caused being outweighed by the need to avoid a greater harm or to further a greater societal interest (i.e.

---

[10] Because of this inclusion, one might prefer to think of both situations as being grounded in excuse, and this is true on both the wrongness and quality of reasoning hypotheses: Dsouza, *Rationale-Based Defences in Criminal Law* (Hart Publishing: 2019) at pp.85ff & 109ff.

[11] CO Finkelstein, "Duress: A Philosophical Account of the Defense in Law" (1995) 37 *Arizona Law Review* 251 at 265.

[12] PH Robinson, "Criminal Law Defenses: A Systematic Analysis" (1982) 82 *Columbia Law Review* 199 at 225-6.

his interpretation of justifications as a form of lesser evils utilitarianism);[13] or exculpation based on conditions which suggest that an actor is not responsible for his conduct (i.e. his interpretation of excuses as recognising varying degrees of incapacity).[14] Thus, Robinson sees circumstances where an actor commits a crime with a loaded gun to their head as representing a loss of control which is sufficiently serious to place the actor in the "realm of abnormality" and hence to distinguish the actor's ability to control his conduct from that of the general population.[15]

This voluntarist language is prevalent in the reported Scottish cases. In *Thomson*, for example, Lord Wheatley stated that a backwards and inferior role in the perpetration of an offence represented one factor to be taken into consideration when determining whether an accused's conduct was "voluntary or coerced".[16] Likewise, in *Moss* the court stated that the act must have dominated the mind of the accused, such that if there is a reasonable legal alternative to disobeying the law, "then the decision to disobey becomes a voluntary one".[17] In *Lord Advocate's Reference* the court stated that necessity required the accused to be "impelled to act as he did"[18] and approved comments by Dickson J in the Canadian Supreme Court case *R v Perka* that in cases of necessity an accused was "remorselessly compelled by normal human instincts" and that such "involuntary conduct should be excused".[19] Even Hume required the accused to have acted "entirely against his will" for a defence of compulsion.[20] A similar approach has been taken in English law in relation to duress by threats.[21] What is particularly odd about the reliance placed on *Perka* in *Lord Advocate's Reference* is the fact that the court seemingly missed Dickson J's broader point

---

[13] *Ibid* at 213. See also PH Robinson, "Competing Theories of Justification: Deeds v. Reasons" in AP Simester & ATH Smith (eds), *Harm and Culpability* (Oxford University Press, 1996) pp.45-70 arguing for an understanding of justifications which maximises the utilitarian ideal of minimising 'net' societal harm. For a critique of the deeds theory generally, see Duff, *op cit*. at 277-80.
[14] *Ibid* at 221.
[15] *Ibid* at 226.
[16] 1983 JC 69 at 77.
[17] 1997 JC 123 at 129.
[18] 2001 JC 143 at 158, para [42].
[19] *Ibid* at 157, para [38], citing *R v Perka* [1984] 2 SCR 232 at 249. Discussed at section 4.4.2.
[20] Hume, i, 52. Indeed, it is probably fair to say that the language of an overborne will stems from the reliance placed by Scottish courts on Hume's treatment of this point.
[21] See, e.g., *R v Hudson* [1971] 2 QB 202 at 206: "the will of the accused has been overborne by threats… so that the commission of the alleged offense was no longer the voluntary act of the accused". An overborne will is a foundational requirement for duress in English law. See, for example: *R v Graham* [1982] 1 WLR 294 at 298; *R v Gotts* [1992] 2 AC 412 at 416. Although cf. Lord Hailsham's comments in *R v Howe* [1987] AC 417 at 428 where he approves of Lord Kilbrandon's comments as the minority in *R v Lynch* [1975] AC 653 at 703 suggesting that an overborne will is nevertheless "a calculated decision". In *R v Safi* [2004] 1 Crim App R 14 the court suggested at [24] that duress by threats and duress of circumstances are both capable of being understood as either the defendant's "choice being overborne" or as "an unwilling choice between two alternatives", adding that the range of circumstances that such defences encompass means that in some cases the references to an overbearing choice may be inappropriate. Thus far, however, it seems that the language of an overborne will is reserved for cases of duress by threats.

about the precise nature of the 'involuntariness' under discussion in necessity type cases. Indeed, Dickson J went on to say that the standard voluntariness requirement "simply refers to the need that the prohibited physical acts must have been under the conscious control of the actor",[22] and thus the type of 'involuntariness' being referred to in cases of necessity (where an actor is clearly in conscious control of their actions) is 'moral involuntariness' (to be discussed in more detail below).

This understanding of voluntariness as pertaining only to 'conscious control' is supported by Aristotle, who considered that actions undertaken out of necessity should be regarded as 'mixed',[23] in the sense that they feature both voluntary and involuntary components, "but are more like voluntary actions; for they are chosen at the time when they are done".[24] Aristotle recognised the difficulty of such 'voluntary actions', in the sense that they are acts that no one would choose in themselves,[25] but concluded that they must be voluntary because they are nevertheless chosen, in contrast to external forces where an agent contributes nothing. To admit the opposite would be absurd says Aristotle, as it would imply that every external thing, good or bad, is constantly compelling us to action.[26] Likewise, Hart states that in situations of coercion an actor has no free or independent choice, and is thus "merely an instrument" for the coercer, but he is "not an instrument in the same sense that he would have been had he been pushed… against a window and broken it: unless he is literally paralysed by fear of the threat, we may believe that [he] could have refused to comply".[27]

Thus, whatever our intuitions are about persons in situations of individual emergency or extreme pressure being less blameworthy, it cannot be based on defects with the voluntariness of an actor's conduct. This is because although external circumstances may influence their decision to break the law, the actor alone is responsible for making the final choice to carry out the offence.[28] We should therefore be very careful when using the

---

[22] *R v Perka* [1984] 2 SCR 232 at 249.
[23] Klimchuk suggests that an action is 'mixed' if under one or more descriptions it is involuntary and under one or more of others it is voluntary: D Klimchuk, "Aristotle on Necessity and Voluntariness" (2002) 19 *History of Philosophy Quarterly* 1 at 6.
[24] Aristotle, *The Nicomachean Ethics*, translation by D Ross (Oxford University Press, 2009) at Book III, 1110a10.
[25] *Ibid* at 1010a15.
[26] *Ibid* at 1110b-15.
[27] HLA Hart, *Punishment and Responsibility*, 2nd edn (Oxford University Press, 2008), p.16. See also at pp.98-9 where Hart is discussing the supposed 'act requirement' and notes that, on the basis of the general doctrine understood in English law, conduct is voluntary if the muscular contraction which initiates the conduct can be said to have been desired or 'willed'.
[28] A Brudner, "A Theory of Necessity" (1987) 7 *Oxford Journal of Legal Studies* 339 at 347: "we must reject… any suggestion that inordinately pressing instincts can negate freedom in the formal sense".

terminology of 'capacity', common parlance in the language of excuse, and it is regrettable that so many of the Scottish authorities should base the rationale of these defences on such shaky ground.

Part of the issue with determining a solid rationale for these situations is that their exculpatory nature appears to rest on dual foundations (and indeed this informs the decision here to recognise two novel concepts instead). Aristotle, for example, noted that in some cases a person might be praised, as where some "great and noble objects" were gained by their conduct, and conversely would be blamed if no such benefit could be realised.[29] This speaks to the utilitarian understanding of conduct which seeks to maximise the societal good, and therefore bases assessments on the value of an act and whether a *net* gain can be said to have derived from it. Aristotle qualifies this point, however, by implying that some actions cannot be compelled, in the sense that "we ought rather to face death".[30] In contrast, however, Aristotle noted that some acts were not worthy of praise but nevertheless we might pardon an actor for being subject to a pressure which could be said to overstrain human nature and which no one could be expected to withstand.[31]

5.2.3 A constrained choice

From the above then it seems reasonable to conclude that, irrespective of whether the conduct is praised or pardoned, such actors are exculpated in relation to the quality of their *conscious* decision to break the law. Specifically, these decisions are exculpatory because while voluntary; they are the product of a constrained *choice*. However, clearly not just any constraint will do – my options for eating might be limited to whatever food I have available in my kitchen, but this fact will not provide any moral or legal mitigation if, unhappy with these options, I decide to steal food from my neighbour. In other words, a choice is not constrained in a meaningful way for situations of individual emergency or extreme pressure (even if the actor genuinely feels constrained) if one of my options is both lawful and reasonable (but is perhaps financially or spiritually damaging or

---

[29] Aristotle, *op cit.* at 1110a20.
[30] *Ibid* at 1110a26. Aristotle gives the example of matricide where Alcmaeon did so under threat of being cursed by his father. Klimchuk suggests that there is some ambiguity here as to whether Aristotle believes such a claim would be "absurd" based on the hollowness of the threat, the seriousness of the crime, or some combination of both. In either case, he argues that 'force' carries some normative weight – it is not just a statement of fact but also a demonstration of appropriate values: Klimchuk, *op cit*. at 9.
[31] *Ibid* at 1110a20-25.

inconvenient for me).[32] There must therefore be an objective, normative component which operates to determine whether an actor's choice has been constrained in this meaningful sense.[33] This component operates to suggest that a choice will be constrained (sufficient to provide exculpation) if the only lawful option that was open to the actor was to bear the harm, and that option was one that the actor was "entitled not to adopt".[34]

The phrase 'entitled not to adopt' is perhaps misleading, however, as it implies that the 'lawful' option of bearing the harm is in some way an 'active' enterprise on the part of the actor; the reality is that it is usually *inaction* which results in the threatened harm materialising. This can be seen most vividly in the classic 'Trolley Problem',[35] where a train cart hurtles towards a group of five men making repairs further down the track and the cart's brakes have malfunctioned. If no action is taken the five men, unaware of the impending danger, will all die. A bystander[36] is able to see this situation unfolding and can divert the cart so that it will instead travel down a second track, but doing so will result in the death of one man who is working on this separate track.

It is clear from this example that if the bystander chooses to switch the track they will be responsible for killing the one worker, but we might say that this choice was nevertheless permissible because it saved five lives. Irrespective of whether one wishes to rationalise this conclusion on the basis that some 'great and noble object' was gained, or because the pressure of the situation was one that nobody could withstand, our intuitions likely tell us that the bystander should not be criminally liable for homicide (or at least murder) if they choose to switch the track. The constrained choice the bystander faced, illustrated by the 'here and now' of the situation[37] and the fact that they were motivated to avert a greater harm, leads us to the conclusion that they should be treated differently to a person who otherwise intentionally causes death.

However, Alexander has pointed out that in real life situations there will often be more than two options available to actors making difficult choices, serving to complicate

---

[32] See, e.g. MS Moore, *Placing Blame: A Theory of the Criminal Law* (Oxford University Press, 2010), at p.560-1 discussing the normative aspect of 'unfair opportunities'. See also A Wertheimer, *Coercion* (Princeton University Press, 1987) at p.10.
[33] This role is often filled by proportionality requirements in various defences.
[34] Gordon, *Criminal Law*, at para. 13.01.
[35] As popularised in Judith J Thomson, "The Trolley Problem" (1985) 94(6) *The Yale Law Journal* 1395.
[36] The original problem developed by Philippa Foot in 1978 saw the driver of the cart making the choice of whether to switch the tracks – Thomson thinks this problem is easier to solve given that if the driver chooses to do nothing, they will still be 'driving over' the five (i.e. actively killing them), and thus it is better to actively kill one than five: P Foot, "The Problem of Abortion and the Doctrine of Double Effect" in *Virtues and Vices and Other Essays in Moral Philosophy* (Oxford University Press, 2002 (reprint)), pp.19-31.
[37] Thomson, *ibid* at 1400.

the assessment.[38] In the context of lesser evils reasoning – what Aristotle called acting for some 'great and noble object' where exculpation exists for making the 'right' choice – the question then becomes one of whether the actor must choose the *least* evil, or is permitted to follow just a *lesser* evil option. Alexander thus adapts the Trolley Problem so that the choices are between killing one to save five, killing two to save five (there is now a second track that the bystander can divert the cart to), or letting the five die.[39] From an objective standpoint, the *least* evil course of action would be to kill the one thus saving five lives.[40] However, Alexander argues that if the bystander is not motivated to kill the one (perhaps because it is someone they know) then they should still have a defence if they choose to kill the two to save five because, in the absence of affirmative duties,[41] the lesser evils defence is a *permission* and not a mandate. The bystander is under no obligation to help anyone (and certainly will feel disincentivised if they expect to be punished for doing so). Indeed, "if the choice is between five deaths and two, we would prefer that he choose two. And if he is not motivated to save the five by killing the one, but *is* motivated to save the five by killing the two, we should be thankful that he at least has some motivation to save the five."[42]

Likewise, where a person acts under a situation of individual emergency to protect a legitimate interest, the law is not generating any kind of mandate as to which right is paramount to protect: exculpation here is based on a permission derived from a recognition that certain rights may be protected in certain emergency situations. In contrast, where a person acts due to some overwhelming pressure which no one could withstand this issue disappears – if a person is overcome by emotion it seems illogical to expect them to somehow identify and follow the best option despite this defect.[43] Indeed, we might even

---

[38] L Alexander, "Lesser Evils: A Closer Look at the Paradigm Justification" (2005) 24 *Law and Philosophy* 611 at 619.
[39] *Ibid* at 618ff.
[40] Those of a Kantian persuasion may, however, beg to differ on the basis that human lives are incommensurable and should never be used as a means to an end.
[41] Such as the duties owed by a parent to their child, or similarly to someone who has voluntarily assumed the role of carer to another.
[42] Alexander, *op cit*. at 619. Kent Greenawalt reaches a similar, if somewhat pessimistic, conclusion when considering this question within the context of his argument that precise definitions of 'justifications' and 'excuses' are (practically) impossible, pointing out the inherent difficulties of reconciling the varying nuanced moral judgements of society regarding actions with the blunt nature of an act as being 'justified'. For this reason, he argues that it is best to think of justified acts as including the 'morally permissible' acts which are 'less than ideal': K Greenawalt, "The Perplexing Borders of Justification and Excuse" (1984) 84 *Columbia Law Review* 1897 at 1904-6. Duff appears to impliedly concur with these views when endorsing the term 'permissible' for justifications: *Answering for Crime*, *op cit*. at p.266.
[43] On this point, see *United States v Holmes* (1842) 26 F.Cas. 360 at 364: "Where the danger is instantaneous, the mind is too much disturbed… to deliberate upon the method of providing for one's own safety, with the least hurt to an aggressor… 'I see not, therefore, any want of benevolence which can be reasonably charged upon a man… if he takes the most obvious way of preserving himself, though perhaps some other method

think it reasonable for such an actor to miss another *lawful* alternative which bypasses the harm, although there might be good policy reasons (i.e. reasons which extend beyond the individual and their circumstances) for rejecting the recognition of this normative conclusion in law. 'Choice' in these contexts therefore takes on quite a different meaning. All that matters in such circumstances is that the choice was understandable, in the sense that it was a choice we would not blame anyone for making. Setting aside the more problematic question of whether we normatively desire for citizens to actively kill others when the numbers support the action, the present thesis endorses Alexander's general logic concerning alternatives, and thus both situations of individual emergency and extreme pressure do not require that only one potential alternative course of action exists; rather, they assume that only objectively *bad* alternatives exist.[44]

To briefly summarise the above then, both situations of individual emergency and extreme pressure conceive of actors undertaking seemingly voluntary conduct on the basis of a constrained choice where the only lawful option available is to embrace the threatened harm, an option that the actor cannot be expected to endure or is entitled to reject. We have also touched on the fact that a constrained choice is capable of several interpretations – in other words, it appears to be possible to reach the same conclusion (exculpation by constrained choice) via different normative, exculpatory elements. We shall now consider these elements in greater detail.

5.2.4 Choices and desires

One particular strand of the voluntarist literature seeks to determine whether or not there is a 'voluntary act principle' in criminal law and, if so, what this might look like.[45] In essence, it is a more sophisticated understanding of voluntariness which attempts to accommodate the 'hard cases', such as commission by omission, and the kind of situations under scrutiny here.[46] Chiao has, for example, put forward a 'practical agency condition'

---

might have been found out, which would have preserved him as effectually, and have produced less hurt to the aggressor, if he had been calm enough, and had been allowed time enough to deliberate about it.'"

[44] Paley, *op cit.* at 205: "[N]ecessity involves the problem of choice between at least two avenues of conduct". Again, this might be a matter of debate in the context of overwhelming pressure which causes an actor to miss a lawful method of escape.

[45] See, e.g. Hart, *op cit* at pp.90-112.

[46] For example, Klimchuk has attempted to reimagine Aristotle's understanding of voluntariness such that actions taken under necessity or coercion might be described as involuntary on the basis of a contextual interpretation of Aristotle's statements in *Nicomachean Ethics* read in line with other statements made in his *Eudemian Ethics* and *Magna Moralia*: Klimchuk, *op cit*. One issue with Klimchuk's interpretation is that it rests on an assumption that human nature is not 'chosen', in the sense that one never *chooses* to act out of self-preservation. It is not immediately clear that this is the case. It certainly seems that, in at least some valid

for liability, which broadly states that an agent is not responsible for the consequences of their conduct unless they were caused or constituted by the agent's conduct *qua* practically rational agent.[47] Likewise, Klimchuk argues that actions performed under coercion or necessity are excused because the reduction in the actor's capacity to choose can be seen as their autonomy being compromised, and thus their actions 'resist imputation to them': the actor's agency is not implicated in their doings.[48] Botterell has suggested that in situations of necessity or coercion we excuse because we think that, owing to the circumstances, the actor's will is overborne by external factors and is therefore not a will that the actor embraces, an 'embraced will' constituting a necessary condition for criminal liability.[49]

All of the accounts above seek to draw a necessary connection between voluntary, physical actions and an underlying will or desire to complete these actions. If this will is absent, for whatever reason, the conduct may be regarded as involuntary in a *normative* sense.[50] The Canadian Supreme Court in *R v Perka*[51] considered the nature of an actor's conduct in necessity situations and concluded that such actors suffered from a defect in their voluntariness in the sense that their actions were impelled. Recognising the difficulties of understanding the nature of such conduct highlighted by Aristotle, and in contrast to the ambiguous voluntarist language that we see pervade the Scottish judgments, Dickson J argued that while the action itself should be regarded as voluntary, the actor's choice was "no true choice at all" and, echoing the above literature, thus should be regarded as 'morally involuntary' to reflect this fact.[52]

'Moral involuntariness', according to Dickson J, is predicated on Hart's theory of excuses that it is preferable to live in a society where we have the maximum opportunity to choose whether we shall become subject to criminal liability.[53] Correspondingly, it is desirable for the state to treat its citizens as self-actuating, choosing agents. To this end,

---

scenarios, an actor actively chooses to prefer their own interests of self-preservation over other interests, and is entitled to do so.

[47] V Chiao, "Action and Agency in the Criminal Law" (2009) 15 *Legal Theory* 1 at 16.

[48] D Klimchuk, "Moral Innocence, Normative Involuntariness, and Fundamental Justice" (1998) 18 *Criminal Reports* (5th) 96 at 102. This point was cited with approval by the Supreme Court of Canada in *R v Ruzic* [2001] 1 SCR 687 at [46].

[49] A Botterell, "Understanding the Voluntary Act Principle" in F Tanguay-Renaud & J Stribopoulos (eds), *Rethinking Criminal Law Theory: New Canadian Perspectives* (Hart Publishing, 2012), pp.97-115 at 110-1. Botterell refers to this requirement as the 'Willed Conduct Requirement'.

[50] Brudner refers to 'concrete freedom' as being determined by a harmony of inner powers and outward circumstances with the peculiar desires of the individual on the other: *op cit.* at 349.

[51] [1984] 2 SCR 232. Discussed above at section 4.4.2.

[52] *Ibid* at 249. This reasoning was followed in relation to duress cases: see *R v Hibbert* [1995] 2 SCR 973. See also the subsequent judgment of Lebel J in the duress case, *R v Ruzic* [2001] 1 SCR 687 at [44]: "her will is overborne… Her conduct is not, in a realistic way, freely chosen".

[53] Hart, *op cit*. at pp.22-4 and 49.

and to ensure just punishment, Hart argued that actors should possess the normal physical and mental capacities required for abstaining from what the law forbids, and a fair opportunity to exercise those capacities.[54] On the basis of these principles, Fletcher states that an actor should be excused if they lack adequate choice.[55] Dickson J agreed with Fletcher's assessment and considered it unjust to punish violations of the law in circumstances where the actor had no other viable or reasonable choice available: "the act was wrong but it is excused because it was realistically unavoidable".[56] Thus, moral involuntariness is closely bound to the concepts of free will and agency in the sense that their absence represents a 'compulsion of circumstance';[57] a choice is constrained (sufficient for exculpation) if it disallows an actor the opportunity to express their inherent autonomy. The implication is that in such circumstances an actor simply reacts, rather than choosing to respond to the situation, and when this happens their autonomy is sufficiently compromised and it can be said that no choice exists at all.[58] In other words, the actor (re)acts from emotions such as fear, not from choice, when committing such acts.[59]

5.2.5 Constrained choices and constrained wills

The concept of moral involuntariness articulated by the Canadian Supreme Court, as a theory of excuse whereby a person's will is impaired by circumstantial pressure, necessarily presupposes wrongdoing because it recognises that an actor can morally object to the actions they are forced to undertake. Indeed, if the actor's conduct is desirable they have no reason to alienate their will from their action. Thus, while an impaired will can certainly explain many of the situations described at the beginning of this chapter, it is too narrow a concept to explain them all. It cannot explain, for example, situations where an actor breaks the law when it is objectively the *right* thing to do – consider again the Trolley Problem but this time switching the track causes the empty cart to travel along a line which ends with a sink hole; we would undoubtedly prefer to lose the cart than to lose five lives, and we would hope that anyone in the bystander's position would choose to switch the track. Such acts are not wrong, nor are they merely understandable; they are positively

---

[54] *Ibid* at p.152.
[55] GP Fletcher, *Rethinking Criminal Law* (Oxford University Press, 2000) at p.804-5.
[56] *R v Perka* [1984] 2 SCR 232 at 250.
[57] Fletcher, *op cit*. "Rethinking Criminal Law" at 804.
[58] G Williams, "Necessity: Duress of Circumstances or Moral Involuntariness?" (2014) 43 *Common Law World Review* 1 at 8.
[59] *Ibid* at 8-9. For a recent Canadian decision reaffirming the principle of fundamental justice that morally involuntary behaviour should not be penalised in necessity cases, see *R v Nwanebu* 2014 BCCA 387.

desirable. It seems inaccurate to describe the bystander as suffering from an impaired will in this scenario because their choice is likely an easy one, corresponding to their desires.[60] Indeed, Coughlan et al. note that "not all threats legally negate a free, realistic or fair choice, or legally overbear or constrain an accused's will".[61]

Thus, and to return to the concepts introduced at the beginning of this chapter, we must separate the kinds of situation that warrant the exculpation of an actor into two categories to reflect these different understandings of choice: those where an actor's will becomes impaired by the extremely stressful stimulus; and those where the actor's will remains very much intact but circumstances compel a difficult choice. Situations of individual emergency concern instances where the actor is faced with a perceived wrong for which they believe they are entitled to respond in averting or diverting that harm by engaging in conduct which would ordinarily be criminal. In contrast, situations of extreme pressure are those which the Canadian Supreme Court would categorise as involving moral involuntariness, where what is done is clearly wrongful, and instead the source of exculpation comes from the actor themselves and the fact that they do not embrace the wrongful conduct which they nevertheless undertake.[62] In folk psychology terms, we can say that it is the difference between acting out of emotion (i.e. a traditionally understood 'overborne will') versus deliberating about the available options and striving to protect a greater interest (assuming a certain level of reasonableness/understandableness has been reached).[63]

Respecting this division involves recognising different understandings of a constrained choice and volition generally. Wertheimer argues that in the context of the voluntariness principle, there are three senses in which someone can be said to be acting in an involuntary manner.[64] First, we might describe one's *movements* as being involuntary such that there is no will present to associate with the act (as where someone twitches or spasms). Second, we might describe one's *acts* as being involuntary, such that the act proceeds from the actor's will, but their will has been impaired by some internal condition or external pressure. Involuntary *movements* involve the absence of volition, whereas involuntary *acts* involve a defect of volition; they are thus both types of nonvolitional

---

[60] On these 'easy' decisions, see GP Fletcher, "The Individualization of Excusing Conditions" (1974) 47 *Southern California Law Review* 1269 at 1277.
[61] S Coughlan, G Ferguson, D Stuart, B Berger, C Mathen & P Sankoff, "Reform of the Defence of Duress (and Necessity)" (2018) 66 *Criminal Law Quarterly* 229 at 230.
[62] Brudner, *op cit*. at 348.
[63] Greenawalt, *op cit*. at 1912: "Someone who is fully rational can be coerced if the threatened harm is so great that he is no longer 'free' to choose that harm."
[64] Wertheimer, *op cit*. at p.9.

acts.[65] Finally, Wertheimer describes the third sense of involuntariness as a 'constrained volition', whereby an actor is confronted with unwanted alternatives, but is quite capable of making a rational choice among them.[66]

If we transpose this understanding to the conceptual framework presented here, we can say that situations of extreme pressure involve nonvolitional acts due to a constrained will, whereas situations of individual emergency involve a constrained volition, otherwise understood as a constrained choice. This will vs choice dichotomy presents us with a more conceptually appealing distinction between the coercion and necessity defences than is currently found in Scots law, with coercion here representing situations of extreme pressure,[67] and situations of individual emergency providing the normative basis for necessity.[68]

Traditionally, coercion and necessity have been distinguished based on whether the threat originated from a human or natural source respectively.[69] Spain has suggested that the availability of an alternative defendant in the case of coercion where no such person is available in cases of necessity is one factor which points towards the source of the threat as being a relevant distinguishing feature.[70] The fact that human threats can ground pleas of necessity, however, suggests that this basis for a distinction is weak.[71] Likewise, Yeo's contention that the distinction is relevant because a threat from a natural source is easier to avoid than one from a human source wrongly assumes a degree of typicality of circumstances in all cases which simply does not exist.[72] Consider again the hiker who

---

[65] *Ibid*.
[66] *Ibid*.
[67] Given the focus on an impaired will in situations of extreme pressure, one might find the term 'duress' to be more appropriate. Alternatively, Anglophobes might prefer the historical Scottish term 'compulsion'.
[68] Self-defence would also fall under this latter category, with the two defences being distinguished by the identity of the victim as also being the initial aggressor, in the case of self-defence.
[69] *Moss v Howdle* 1997 JC 123 at 127 per Lord Justice-General Rodger, affirming the judgment of Lord Hailsham LC in the English case of *R v Howe* [1987] AC 417 at 429: "There is, of course, an obvious distinction between duress and necessity as potential defences; duress arises from the wrongful threats or violence of another human being and necessity arises from any other objective dangers threatening the accused." From legal literature see, for example: MD Bayles, "Reconceptualising Necessity and Duress" (1987) 33 *Wayne Law Review* 1191 at 1191; J Dressler, "Exegesis of the law of Duress: Justifying the Excuse and Searching for its Proper Limits" (1989) 62 *Southern California Law Review* 1331 at 1348; Wertheimer, *op cit* at p.146; G Williams, *Criminal Law: The General Part* (2nd edn, 1961) at p.757; Law Reform Commission of Ireland, Consultation Paper: "Duress and Necessity" (Ireland: LRC CP 39, 2006), at para 1.01 (available at www.lawreform.ie/_fileupload/consultation%20papers/Duress%20and%20Necessity%20CP.pdf).
[70] E Spain, *The Role of Emotions in Criminal Law Defences: Duress, Necessity and Lesser Evils* (Cambridge University Press, 2011) at p.3.
[71] See, e.g. *Tudhope v Grubb* 1983 SCCR 350; *R v Willer* (1986) 83 Cr App R 225.
[72] S Yeo, *Compulsion in the Criminal Law* (Law Book Co, 1990), at pp.26-7.

breaks into a cabin to escape a snowstorm: it is at least arguable here that the natural threat is *harder* to avoid than some human variants.

Chalmers and Leverick have alternatively suggested that the distinction is based on whether the choice is one dictated by another (coercion), or is made reluctantly by the accused themselves (necessity).[73] It is submitted that the distinction proposed here, focusing on whether the situation represented a constrained choice or constrained will, is broadly compatible with this understanding – at least insofar as Chalmers and Leverick's theory recognises the impact that an impaired will can have on the making of choices – but places a greater emphasis on the accused's agency.[74]

Finally, Spain has suggested that a relevant distinction might be found in the general division of exculpatory defences, with necessity being typically understood by commentators as a justification and coercion as an excuse.[75] Through a thorough analysis of the necessity and duress defences in both English and Irish law, she concludes that duress should be regarded as an excusatory defence where threats impel fear and this fear constitutes a reasonable emotional reaction to the circumstances.[76] In contrast, necessity would operate in a residual, justificatory capacity as a utilitarian-styled lesser evils defence, where the accused acts to minimise harm irrespective of their emotional state.[77] She therefore identifies two sites of division between the defences – one which has (broadly) been recognised here (constrained choice vs constrained will), and another based on the justification/excuse distinction.[78]

While Spain's theory has proved influential on this thesis, I am unconvinced by this second basis for division, and as aforementioned would rather avoid the justification/excuse distinction here as it tends to be conceptually unhelpful and unnecessary. A variation of the Trolley Problem may help to explain my doubts. Above I

---

[73] Chalmers & Leverick, *Defences*, para. 5.02. For a similar view see R Nozick, *Philosophical Explanations* (Clarendon Press, 1981), at p.520, who argues that whereas one exercises their own will in response to natural threats, when the threats are 'man-made' they tend to involve another's intentions such that it makes the action not truly one's own.

[74] Where we would presumably differ, theoretically speaking and in terms of categorisation, is where a person suffers from an impaired will which cannot be attributed to a coercer. To return to the Trolley Example, imagine the trolley will hit the accused's child unless they switch the track to one which has two workers on it: I would classify this as a situation of extreme pressure and hence coercion, whereas Chalmers and Leverick would likely consider this to be an example of necessity.

[75] Spain, *op cit*. at p.4. See also PH Robinson, *Criminal Law Defenses* (1984), §124(e)(1).

[76] *Ibid* at p.263.

[77] *Ibid* at p.283.

[78] J Horder recognises a similar distinction, save that his understanding of duress/coercion is of personal sacrifice, and whether it is reasonable for society to expect persons to make such a sacrifice in the particular circumstances: J Horder, "Self-Defence, Necessity and Duress: Understanding the Relationship (1998) 11 *Canadian Journal of Law and Jurisprudence* 143. Discusses more fully below at 9.4.

gave a variation of the problem where the second track led to a sinkhole as providing an example of a situation which might warrant exculpation despite the absence of a constrained will or wrongful act. Consider now a third variation, where the bystander believes that pulling the lever will divert the trolley hurtling towards the five workers onto the sinkhole track and so pulls it, but the lever instead sends the trolley to a third, previously hidden track which kills six workers making repairs there. Spain's conception of necessity as a lesser evils defence, requiring a minimisation of harm, would leave the bystander with no defence, nor could they rely on her conception of duress because they had no impaired will.[79] Under my understanding of necessity, understood as a situation of individual emergency, the bystander would have a defence because their actions were understandable, and anyone in their position would have done the same.[80] Likewise, if the actor turned out to be correct about pulling the lever and saved the five lives by destroying the trolley, we can equally say that their actions (destruction of property) were understandable and thus deserving of a necessity defence. We can of course say that such actions were laudable, but I contend that this extra step is unnecessary for the purposes of determining criminal liability.

Thus, a division based solely on a constrained choice and constrained will provides the most normatively pleasing structure from which to form these reactive defences. To summarise, I contend that the types of conduct currently covered by necessity and coercion in Scots law should instead be divided based on whether the facts suggest a situation of individual emergency, or a situation of extreme pressure. Situations of individual emergency are those where an actor aims to bring about a positive result in the face of bad alternatives and should be assessed by the presence of a constrained choice. It thus shares a similar evaluative basis as the current necessity and coercion defences, but a different factual nexus. In contrast, and more controversially, situations of extreme pressure are those where an actor reacts out of emotion to protect their interest(s), and is assessed by the presence of a constrained will – this idea of an 'overborne will' which substantially affects an accused's reason for action.

Having established this distinction between situations of individual emergency and extreme pressure and their relationship with constrained choice/will, more needs to be said about the precise contours of exculpating someone on the more elusive basis that they suffered from a constrained will. This is because a constrained choice conforms more

[79] Horder's understanding of the division leads to a similar conclusion: see his treatment of necessity in *ibid* "Understanding the Relationship", *op cit*. at 155ff.
[80] See section 5.2.1 above. Admittedly it is unclear whether Spain is an act or rule utilitarian in this respect.

closely to the current understanding of how necessity and coercion are assessed (i.e. the more natural interpretation of choice) and is something that is objectively verifiable. In that sense it is uncontroversial. In contrast, how and when is a will constrained such that it warrants a negation of criminal liability? Is a subjective feeling enough to ground a defence? What does an emotional reaction look like and are all emotional reactions equal? We have already touched on the fact that not just any constraint will do, and thus a deeper understanding of constrained will should be established, in terms comparable to a constrained choice, such that the division is not just normatively pleasing, but practically workable. The remainder of the second part of this thesis is devoted to further define this concept of a constrained will by its severance from the concept of choice. The next section shall thus further defend a normative distinction between choice and will, and examine the precise contours of this relationship, before embarking on an extensive analysis of emotions in criminal law defences and their place in situations of extreme pressure.

5.3 Understanding the Vitiation of Blame in Situations of Extreme Pressure: Choice, Capacity and Character

In this section I wish to further explore the rationale for the exculpation of persons in situations of extreme pressure, where it can be said that they acted with a constrained will. Specifically, I wish to examine the relationship between will and choice in the context of determining criminal culpability in stressful circumstances, and to challenge the view that a lack of choice is equal to a lack of will. As we saw above, the current understanding of reactive defences sees a lack of voluntariness as negating the choice made. Intuitively, however, it seems like something more, or at least something else, is going on when we say that someone is exculpated because of a constrained will. In contrast, in situations of individual emergency where the actor makes a rational decision based on the limited options before them, the terminology of choice seems appropriate.

Further, exploring this relationship is an important step in solidifying the argument introduced above that these situations are better distinguished on the basis that each defence serves to represent a different sense of constraint. In Hartian terms, it might be said that this section will explore the difference between a lack of capacity caused by being emotionally overwhelmed, versus a lack of fair opportunity owing to emergency circumstances. Through an analysis of constrained will and its relationship with choice it

will be suggested that there is a moral distinction between the two actors, one which can be given appropriate recognition through a proper delineation of reactive defences.[81]

It may be tempting to suggest that, as both the overwhelmed and emergency actors can at least be said to suffer from some degree of constrained choice, a distinction is unnecessary as they both merit exculpation on the basis that they lacked a fair opportunity to conform to the law. However, not only is it conceptually unappealing to describe an overwhelmed actor as having 'lacked a fair opportunity', but it is also descriptively lacking: does not the fear of one's own injury or death warrant exculpation in its own right? It is perhaps even more obtuse to suggest that a person who acts to do the right thing in an emergency was 'emotional' or lacked capacity. There is nevertheless a tendency for legal systems to conflate lack of capacity and lack of opportunity in exculpation. I argue that a normatively defensible position must recognise and appropriately distinguish between situations of extreme pressure and circumstances of individual emergency.

5.3.1 Choice: capacity versus opportunity

As a purely descriptive account, we might say that the distinction between an actor who suffers from an 'overborne will' on the one hand, and one who acts rationally in the face of an emergency on the other, is the presence and psychological impact of strong emotions. Moore locates this distinction in his understanding of the orthodox choice theory of culpability inspired by Hart – we can say that the emotionally overwhelmed actor suffers from a defective *capacity*, whereas the emergency actor suffers from a lack of fair *opportunity*.[82] In necessity and coercion cases, the primary emotion will usually be fear. In this sense, we might say that situations of extreme pressure are analogous to the defence of provocation which also finds its (partially) exculpatory element in strong emotion, namely anger.

However, the incorporation of emotional reactions in law is a difficult task. These difficulties can be highlighted by comparing the Canadian approach to necessity (moral involuntariness) with the English defence duress of circumstances. The reader will recall

---

[81] On degrees of moral blame for such actors, see e.g., JM Fischer, "Responsiveness and Moral Responsibility" in F Schoeman (ed), *Responsibility, Character, and the Emotions* (Cambridge University Press, 1987) pp.81-106 at p.83, fn. 3 who argues that there is a difference in the moral responsibility of each actor (the stoic being morally responsible whereas the emotionally overwhelmed is not), but perhaps not in their (lack of) blameworthiness. I do not attempt here to demonstrate anything more than that such a distinction exists.

[82] Moore, *op cit*. at p.554.

from chapter three that in order for a defendant to claim duress of circumstances, it must be shown that a sober person of reasonable firmness would have reacted in a similar way.[83] Williams has argued that this test wrongly ignores the defendant's emotional state in assessing their reaction to the danger that they were placed in.[84] One might genuinely be overwhelmed by the threat but nevertheless have no defence. There is thus an assumption in English law that individuals should possess a certain level of courage, and it has even been suggested that the standard is that of a hero[85] – one that the average person might never reach. In contrast, Berger warns against moving too far in the opposite direction, highlighting that the moral involuntariness concept places such an emphasis on the supposed involuntary nature of emotions and their effect on an actor's capacity that it serves to withdraw any kind of judgment based on normative critical reflection, thus veiling the normative foundations of the criminal law and potentially endorsing problematic emotional responses to pressurised situations.[86]

A balance must therefore be struck when developing a defences framework which can accommodate both situations of extreme pressure and individual emergency. This balance involves reassessing our understanding of theories of culpability, and what it means to have a 'choice' in the context of constrained will. At a theoretical level, the exculpation of both the emotionally overwhelmed and emergency actors represents an important definitional challenge to reactive defences, because they must be able to accommodate different understandings of a 'constrained choice' as understood by the choice theory and the courts. Under the current construction of these defences, if a defence purports to exonerate on the basis that such situations impair an actor's capacity to choose to some degree, it must do so in a way which can accommodate our intuitions that both the emotionally overwhelmed and emergency actors should be exculpated in at least some circumstances, even when their actions are wrongful.[87] In other words, 'capacity' here cannot refer to a state (just short) of automatism, where the actor has lost the ability to reason (as it does in other defences such as mental defect or involuntary intoxication), as then our emergency actors who break the law in understandable conditions (which

---

[83] See section 3.3.1; *R v Martin* [1989] 1 All ER 652.
[84] Williams, *op cit.* at 5.
[85] *Ibid* at 6. See also *R v Howe* [1987] AC 417 at 432, per Lord Hailsham, in relation to the related defence of duress by threats. Cf. *R v Ruzic* [2001] 1 SCR 687 at [40] per Lebel J: "The law is designed for the common man, not for a community of saints or heroes". Cf. Fletcher discussing the German criminal law approach: "the German law's cultivating excusing conditions is tied to indifference toward fictitious standards of exemplary men". Fletcher, *op cit*. "Individualization of Excusing Conditions", at 1290.
[86] BL Berger, "Emotions and the Veil of Voluntarism: The Loss of Judgment in Canadian Criminal Defences" (2006) 51 *McGill Law Journal* 99 at 103.
[87] RA Duff, "Choice, Character, and Criminal Liability" (1993) 12 *Law and Philosophy* 345 at 356.

nevertheless fall short of permissible conduct) would be left outside of the framework, and it is clear that in some cases such a conclusion would be incorrect. Indeed, in many cases the actor is excused precisely for *expressing* that ability to reason.

Conversely, if the basis of such defences is grounded in a lack of fair opportunities, understood as a lack of a realistic choice, we must be mindful of what it means to 'have no choice' in situations of extreme pressure, and whether an actor's capacity can and should play a role in determining the existence of choice. Freedom and choice are, after all, not terms free of ambiguity. Indeed, Brudner has highlighted that while other excuses such as mistake, involuntary intoxication, and insanity/mental disorder defences can be said to vitiate freedom in a formal sense, such that they deny attributability of the act[88] to the actor and thus excuse on the basis that the actor "did not choose the consequence he caused",[89] it becomes very difficult to delineate what is meant by a more 'robust' sense of freedom which can accommodate cases where the actor concedes attributability but denies that the reasonable man of ordinary firmness would have acted differently.[90] In essence, just what counts as a valid expression of the actor's free choice? To illustrate this point, Brudner asks us to compare the classic 'kill or be killed' scenario with the 'rotten social background' (RSB) example. In the RSB example, external, structural factors outwith the actor's control – such as the circumstances of their upbringing, natural endowments, and a depressed economy (perhaps caused by a global pandemic!) – have operated to force the actor to choose between relying on social welfare or turning to a life of crime.[91]

If the actor should choose to commit crime to survive, Brudner thinks that voluntarism and choice theorists will have a hard time explaining why this actor is any different from the person who kills out of necessity. In both scenarios it can be said that none of the options are attractive in themselves and that the actor would avoid picking any of them if they could. Thus, if our intuitions tell us that only the actor in life-or-death situations should be excused, it must be on the basis that, in contrast to our intuitions regarding the RSB actor, we understand and are sympathetic to the actor's failure to overcome their aversion to death.[92] In other words, something other than 'choice' is

---

[88] Of course, in one sense the only true denial of attributability is a denial by the agent that they committed a crime at all, i.e. that there was no *actus reus*. In that sense, and to borrow from Hart, attributability here should be treated more in line with legal liability responsibility, rather than causal responsibility: Hart, *op cit*. at p.211ff.
[89] Although cf. Duff, "Choice, Character and Criminal Liability", *op cit*. suggesting that in some cases the focus may be on the mentally disordered agent's inability to appreciate the normative aspects of his actions.
[90] Brudner, *op cit*. at 345-50.
[91] *Ibid* at 348.
[92] *Ibid*.

grounding our intuitions about exculpation here. From another perspective, Horder has pointed out that choice necessarily presupposes a comparison between alternative courses of action – with this in mind it becomes hard to rationalise why a person who impulsively kills another should not be entitled to exculpation on the basis that their action lacked any real sense of choice in considering alternative courses of action.[93] Likewise, it makes very little sense to talk of the negligent wrongdoer as having, in any meaningful sense, 'chosen' to do wrong.[94] On this basis, Horder accuses the choice theory of being reliant on a broader theory of capacity to explain culpability.[95]

Equally, Brudner challenges an understanding of necessity based in terms of a denial of attributability (i.e. an orthodox understanding of capacity) at a conceptual level, specifically questioning the compatibility of the 'robust' or 'concrete' view of freedom (i.e. the idea that an actor must not only choose the consequence, but also the content of that choice must be one which they can affirm as their own) with the idea of wrongdoing being understood as purposive acts of the self, in contradiction to the law.[96] On this construction of wrongdoing, defences like mistake, involuntary intoxication, and mental disorder exculpate on the basis that there is a denial of the wrongdoing in question being attributable to a conscious, purposive act of the self. In other words, there is no rational cognitive process from which one can link the act to the actor, or a lack of capacity. Persons in situations of extreme pressure, however, cannot (always) make such claims because they positively choose their consequences, and it makes no difference that the choice does not correspond to their ultimate desires, because wrongdoing only requires the purposive act.

It is for this reason, says Brudner, that the law treats motives as irrelevant to culpability – they are all motives of an individual ego.[97] In terms of the law's authority, it does not matter whether an unlawful consequence was desired or unwillingly accepted as a necessary means of satisfying some other desire: in both situations the law becomes unacceptably subordinated to the primacy of the person, who is themselves the subject of desire.[98] In other words, exculpation on the basis of a denial of 'concrete' freedom necessarily involves subordinating the law to individual desires in a way that is

---

[93] J Horder, "Criminal Culpability: The Possibility of a General Theory" (1993) 12 *Law and Philosophy* 193 at 201.
[94] *Ibid* at 199.
[95] *Ibid* at 199-200.
[96] Brudner, *op cit*. at 349.
[97] See also MD Bayles, "Character, Purpose, and Criminal Responsibility" (1982) 1 *Law and Philosophy* 5 at 19, stating that the law is concerned only with more specific dispositions which produce social harms.
[98] Brudner, *op cit*. at 349.

inconsistent with the accepted understanding of excusing conditions (which negate the attribution of wrongdoing to the actor). On this logic, Brudner is sceptical that *any* theory of excuse that utilises the notion of wrongdoing (purposive acts in contradiction to the law) can integrate reactive defences like necessity or coercion with defences like mistake, involuntary intoxication, and mental disorder.[99]

Moore, an ardent proponent of the choice theory of responsibility, responds to these types of arguments (and Horder specifically) by arguing that when a choice theorist speaks of exculpating when an actor 'could not have done otherwise', this statement should be treated as elliptical for 'they could not have done otherwise if they had chosen to'.[100] Thus, what makes the intentional or reckless wrongdoer so culpable (i.e. more culpable than inadvertent wrongdoers) is not bare unexercised capacity, but the way such capacity goes unexercised: the actor has *chosen* not to exercise it.[101] This is to be distinguished from the kind of 'unexercised capacity' referred to in negligence cases, where a person is unaware of the wrongness of their action, there being a 'fundamental' difference between an actor who knows that they are doing the sort of action that is (likely to be) wrongful and chooses to do it anyway, and an actor who *ex post* learns that their actions were wrongful.[102]

Moore recognises this inadvertent risk creation as a separate kind of culpability which he sees operating in tandem with culpability for chosen wrongdoing. He concedes that he has no major objection to viewing choice as part of a broader capacity theory as Horder claims it to be, but suggests that in doing so one would have to find some other way of recognising the distinction between inadvertent and chosen wrongdoing.[103] As to the challenge about how to define a robust sense of freedom sufficient to define when an opportunity can be regarded as 'fair', Moore is less persuasive. He suggests merely that the assessment is an intuitive one, and thus the reason why a heroin addict who steals money to fuel their addiction has no defence is because to say otherwise would be "morally implausible".[104]

Now it is clear that persons who act in situations of extreme pressure are not, in any sense, inadvertent risk creators by definition; the claims 'I was forced' or 'I had no choice'

---

[99] *Ibid*.
[100] Moore, *op cit*. at 553.
[101] *Ibid* at 590.
[102] *Ibid* at 591-2. This argument is made in direct response to Horder questioning the choice theory's ability to explain criminal liability for negligence, *op cit*. at 199.
[103] *Ibid* at 592.
[104] *Ibid* at 555. Cf. Hirstein, Sifferd & Fagan, *op cit*. at 124 suggesting that an unwilling addict does not want or desire heroin, even if they go to considerable lengths to obtain it.

imply the necessary cognisance to take us out of that territory. Nevertheless, the argument is worth mentioning because it does imply that a theory which is broader than just tracking the presence of a meaningful choice might underly culpability generally, which in turn might help us explain why we seek to exculpate both the emergency and overwhelmed actors. Indeed, Horder's claim that choice, understood as a decision between alternatives, is too narrow to provide a complete theory of culpability is compelling, and thus I will take Moore at his word about the 'harmlessness'[105] of referring instead to a broader, complex capacity theory, safe in the knowledge that I do not here have to find some other basis for recognising the culpability distinction between inadvertent and choosing wrongdoers.

I do wish to highlight (again), however, that the issue of differentiating between types of culpability on the choice theory is perhaps most acute in relation to situations of extreme pressure, owing to the hybrid nature of the 'choice' that such actors make.[106] Typically in the criminal law we *excuse* actors when their action is wrong but the actor is not blameworthy; we can here refer back to Brudner's point about the other recognised excuses denying attributability for the act as explaining why such actors are seen as 'blameless'. However, an understanding of blame, wrongdoing and culpability which is predicated on choice alone seems ill-equipped to deal with situations where the actor has, despite their emotions and desires, positively made a clear choice: the actor is left in a quasi-blameworthy state where many think that they are not entirely blameless.[107] This issue is compounded by Moore's admission that internal factors like emotions cannot be said to incapacitate our choices,[108] and thus even some of the more 'orthodox' examples of culpability being vitiated by a lack of capacity need to be reassessed.

---

[105] *Ibid* at 592.

[106] Although as Horder rightly points out this issue is a general one for a unified choice/capacity theory: "[B]oth the murderer and the involuntary manslaughterer miss a fair opportunity to avoid wrongdoing", Horder, *op cit.* at 203.

[107] Williams, *op cit*. at 12; Brudner, *op cit.* at 351; Horder, *op cit*. at 201 and 203; *R v Ruzic* [2001] 1 SCR 687 at [39] per Lebel J: "in my opinion, conduct that is morally involuntary is not always intrinsically free of blame" (see also paras [40]-[41]). It should be noted that Lebel J seems to operate under the premise that where the elements of an offence have been made out, the accused should no longer be considered 'blameless' – it is unclear then whether Lebel J would consider a person acting in self-defence and other justificatory defences as 'blameless' either: at para [41]. See also GT Trotter, "Necessity and Death: Lessons from *Latimer* and the Case of the Conjoined Twins" (2003) 40 *Alberta Law Review* 817. Trotter argues that, on the basis of the s17 Canadian Criminal Code list of excluded defences to the statutory duress defence, we can surmise that an actor cannot commit serious offences of intentional violence when confronted with immediate threats of death or serious injury, since "the accused has no greater claim to physical integrity than the victim": see at 834-6.

[108] Moore, *op cit*. at 559. See also J Sabini & M Silver, "Emotions, Responsibility and Character" in F Schoeman (ed), *Responsibility, Character, and the Emotions* (Cambridge University Press, 1987) pp.164-75 at 168. See also chapter six generally below.

5.3.2 Fleshing out the normative element: the character theory

Grounding a reactive defences framework purely on a complex capacity theory then, seems unable to exculpate both of our hypothetical actors.[109] On the basis of the criticisms above, while it might seem logical enough to say that the emotionally overwhelmed actor lacks the relevant capacity (although the suggestion that emotions are reason-responsive might suggest otherwise[110]), it becomes harder to justify the emergency actor's lack of fair opportunities without descending into a reductive argument about what counts as 'freedom', and to what extent innate environmental factors might influence how far an opportunity should be regarded as 'fair'. Some commentators, critical of the choice/capacity theories, suggest that exculpation is better explained by the character theory of culpability, whereby judging culpability is determined by focusing on the character traits manifested by actors when they commit offences.[111] Thus, and to put this theory into context, the argument is that the actions of a person who commits a crime in circumstances warranting a reactive defence do not represent who that person really is; they are 'out of character', and thus not deserving of blame. I shall explore this theory here in slightly more detail before moving on to discuss the view that both character and choice theories can, depending on the circumstances, be said to influence exculpation in a defences framework which accommodates both situations of extreme pressure and individual emergency.

As aforementioned, the character theory suggests that when a person claims exculpation, they do so on the broad basis that the act was 'out of character'. In other words, blame and punishment are not assigned directly for acts, but rather for the manifestation of particular character traits. Bayles tells us that character traits are to be understood here as any socially desirable or undesirable disposition, and that acts may or may not indicate such traits.[112] Moore describes this as a 'filtering function' whereby we are only *prima facie* responsible for acts until we determine whether such acts were truly expressive of our characters.[113] If an act does demonstrate an undesirable character trait

---

[109] Duff argues that choice is neither necessary nor sufficient for criminal liability: "Choice, Character, and Criminal Liability", *op cit*. at 364.
[110] See chapter seven below.
[111] See, e.g. MD Bayles, *op cit*. For other variations of the character theory, see: G Fletcher, *Rethinking Criminal Law*, *op cit.* at p800; N Lacey, *State Punishment* (London, Routledge; 1988) at p.65ff; and Horder, *op cit.* at 204ff (agreeing with Bayles' premise for the theory). For a detailed criticism, see Moore, *op cit*. at 562-74.
[112] Bayles, *ibid* at 7.
[113] Moore, *op cit.* at 563. Presumably the choice theory does not 'filter' responsibility in this way because a lack of choice presupposes a denial of attributability in the first place, pre-empting any *prima facie* responsibility, but see the discussion about the validity of this point re Brudner above at 5.3.1 in relation to defences like necessity and coercion.

(i.e. disposition), then blame and punishment may be appropriate. Conversely, where an act does not demonstrate such a disposition (e.g. where someone kills accidentally), then blame is inappropriate (although steps may be taken to prevent such accidents in the future).[114]

Single acts do not necessarily indicate dispositions, but Bayles claims that such dispositions can be inferred from other aspects of criminal law which can serve to corroborate the implications provided by an act, such as the presence of the necessary *mens rea*.[115] Character theorists use this line of logic to suggest that their theory can better accommodate negligence liability than the choice theory, on the basis that negligent actors demonstrate an undesirable character trait worthy of blame (i.e. an unacceptable disregard for others' interests and the law).[116] The difference in culpability between various mental states is accounted for on the grounds that intentional harm is worse in terms of what it shows about an actor's moral character than causing harm negligently. A similar point is made about using character theory to explain attempts, mistakes and mental defects as being situations where the inference of character traits likely to produce criminal harm can justify blame in their presence, or exonerate in their absence.[117]

Indeed, for legal blame and punishment it is insufficient that an actor's intention, knowledge, recklessness or negligence indicates an undesirable trait alone; it must also be a trait which is likely to produce criminal harm.[118] The quality of the action is therefore judged by whether causing harm was the product of something good (on the whole), understandable, indifferent, or bad in the actor's character.[119] Bayles suggests that the point of punishment is to deter and punish bad character,[120] but Horder argues that this view is misguided as it wrongly elevates the character theory from one concerning culpability to one determining liability.[121] To this end, Horder suggests that the character theory is only defensible if grounded in a broader theory of criminal liability which is shaped by some variation of the harm principle where people are not held criminally liable unless they cause harm.

---

[114] Bayles, *op cit.* at 7.
[115] *Ibid* at 8, and at 10: "The behaviour plus the attitude evidence an undesirable disposition." Duff argues that a criminal act "provides the safest evidence of the defective character-trait": see "Choice, Character, and Criminal Liability", *op cit*. at 371.
[116] Bayles, *ibid* at 10.
[117] See *ibid* generally.
[118] Bayles, *op cit*. at 13.
[119] Horder, *op cit*. at 205.
[120] Bayles, *op cit*. at 7.
[121] Horder, *op cit*. at 206.

Lacey has further suggested that, to be culpable, the crime must be attributable to a *settled* disposition, such that if the crime were the result of some uncharacteristic 'slip' then it was not a product of the actor's character.[122] Character is thus (in)formed by verification. Horder, however, argues that such a requirement cannot be correct as it would then render character theories unable to explain liability for negligently caused harm, where the negligence took the form of an uncharacteristic slip.[123] Likewise, and from the other end of the severity spectrum, Moore has noted that the requirement of a settled disposition would make even the most grave harm which was intentionally caused not culpable if it resulted from such a slip.[124] As a result, Horder suggests that the character theory should instead point to an idealised conception of an agent with good character when evaluating an actor's conduct,[125] a position which does lend some credibility to objective tests like the 'reasonable person' test found in duress of circumstances. Duff has similarly suggested that when the character theory exculpates, it does so on the basis that the action expressed no character at all (thus sidestepping the settled disposition requirement completely).[126]

Owing in part to his rejection of any verificationist understanding of character theory (i.e. the problems associated with establishing a settled character sufficient for blame), Moore locates the relationship between actions and character as being evidentiary in nature, in the sense that actions can be seen as evidence of a character trait which caused it, but only if there is some general connection between the class of events that includes the effect and the class of events that includes the cause (i.e. requirements of causality and typicality).[127] We can imagine a Venn diagram where our two inputs are the character trait and action in question and they overlap considerably. In this sense the connection is discriminatory, accommodating for the fact that it is possible for a character trait to cause any number of behaviours on a given occasion, but focusing on only those behaviours that the character trait *typically* causes for the purposes of establishing whether the action was in character.

Many moderate accounts of the character theory thus tie this notion of character as explaining the rationale for culpability with the presence of an act which serves in turn to

---

[122] Lacey, *op cit*. at pp.65-6. See also Duff's discussion of this point as it relates to the implications for the purposes of punishment: "Choice, Character, and Criminal Liability", *op cit*. at 364-70.
[123] Horder, *op cit*. at 207.
[124] Moore, *op cit.* at 578-82 considering the case of Richard Herrin and Bonnie Garland.
[125] Horder, *op cit.* at 207.
[126] Duff, "Choice, Character, and Criminal Liability", *op cit*. at 378.
[127] Moore, *op cit*. at 573.

demonstrate the presence (or not) of this culpable character.[128] We saw above that Bayles considered acts as providing evidence of the dispositions relevant to the law. For Horder, the presence of an act which breaches the harm principle allows us to evaluate what an idealised character, one who respects the law and others' interests, would do in the situation.[129] For Duff, harmful acts are to be seen as a way of appropriately limiting the otherwise overbroad examination encouraged by an assessment of one's moral character.[130] Both Duff and Horder thus conclude that, on the basis of either the harm principle or limiting the scope of the moral assessment the law should be concerned with, this relationship between acts and character is a necessary condition of a liberal conception of the state.[131]

5.3.3 Theories of culpability and emotions

The implications of the character theory are thus that both the emergency and overwhelmed actors should be exculpated where it can be said that the harmful actions undertaken were not representative of any settled character, one which is idealised and assumed from the circumstances of the particular case. If the action belies an underlying negative character trait which is likely to produce harm, then there shall be no exculpation irrespective of whether the actor was genuinely overwhelmed (although we might then seek to (partially) exculpate the actor's behaviour on the basis of a temporary mental defect or diminished responsibility). Specifically, the focus on character aims at levelling the playing field for both of our hypothetical actors by focusing on only one aspect of mental fortitude – whether the traits that can be said to affect fortitude are those which are likely to cause harm. In this way, overwhelming emotions which affect mental fortitude will only exculpate if they demonstrate a character trait which is unlikely to cause harm. This is so even if the actor was genuinely overwhelmed. If such traits would typically cause harm, the actor cannot rely on them for exculpation. Likewise, our emergency actor can similarly point to aspects of character such as resiliency and proactivity when claiming that their actions, while wrongful, were those of a character that the law should not seek to blame or punish.

---

[128] Duff, "Choice, Character, and Criminal Liability", *op cit*. at 367.
[129] Similarly, see J Gardner, "The Gist of Excuses" (1998) 1 *Buffalo Criminal Law Review* 575, arguing that excuses are given to those who "lived up to our expectations", at 578.
[130] Duff, "Choice, Character, and Criminal Liability", *op cit*. at 368.
[131] *Ibid*; Horder, *op cit*. at 206.

Outlining the basic tenets of the character theory here has served to offer a strong challenge to the orthodox choice theory which might otherwise ground the reactive defences framework. Specifically, the comparison drawn has helped to isolate what I believe to be the crux of the issue: how the law should understand emotions and their impact in situations of extreme pressure. We saw above that Moore, in treating emotions as responsive to reason, was forced to categorise what I call situations of extreme pressure as being those where the actor experienced a lack of opportunity. Likewise, under the character theory it seems to make no difference that a person was genuinely overwhelmed by emotion if the action taken would nevertheless demonstrate harmful character traits. In this sense, I am less concerned with outlining the challenges to the character theory as an exclusive theory of culpability: I am only interested in what the theory can potentially say about exculpation in situations of extreme pressure. To this end, I do not here commit to either of the choice/capacity or character theories as an exclusive theory of culpability in the criminal law. My intuitions suggest that character theories better explain situations of extreme pressure, but only insofar as they can better accommodate the position that strong emotions should not be understood to vitiate capacity, but may nevertheless affect liability. In any case, there have been strong arguments to suggest that there is no unifying theory of culpability, and it is interesting to explore how this position might influence our understanding of any reactive defences framework.

Indeed, many scholars suggest that capacity and character can be said to intersect with other elements in determining an actor's culpability. Tadros argues that there is no single conceptual foundation for excuses and thus no reason to restrict the types of excuses which might be regarded as acceptable in the criminal law.[132] Horder adopts a stricter stance in arguing that criminal culpability can be understood through a "patterned mixture" of the capacity, character and agency theories which correspond to special cultural and moral values.[133] In the case of defences like necessity or coercion, the fear of imminent harm or of the use of force is accorded a special cultural and moral significance, such that offences committed in these situations are seen as reasonable expressions of virtues associated with understandable emotions, which are valued aspects of an idealised human character.[134] Duff, in contrast, doubts the ability to distinguish between choice and character in assessing criminal liability, noting that discussions of choice inevitably lead

[132] V Tadros, "The Characters of Excuse" (2001) 21 *Oxford Journal of Legal Studies* 495.
[133] J Horder, "Criminal Culpability: The Possibility of a General Theory", *op cit*: See generally, quotation at 214.
[134] *Ibid* at 209.

back to an assessment of character, and assessments of character invariably lead to discussions about the character of an act.[135] Likewise, and as mentioned above, he argues that the character theory exculpates on the basis that no character (worthy of censure) has been demonstrated by the particular action.[136] On this basis, he claims that the cases where a character theorist would argue an action was 'out of character' (i.e. demonstrate no character), are the same as those where a choice theorist would argue that it was not freely chosen.[137]

Thus, situations of extreme pressure and individual emergency can either be regarded as corresponding to a singular theory (such as the character theory under Horder's tripartite understanding of culpability), or they can be understood under Duff's conception of a mixed theory which synthesises the character and choice theories on the basis that the considerations which would lead to a lack of opportunity under a choice theory, are the same considerations which would lead to an inability to infer harmful character traits under the character theory. Duff's synthesis proposal is normatively appealing because it captures our feelings that at least part of the character theorists' claims about an idealised character, or indeed a refusal to impute any kind of character to a person acting in a situation of extreme pressure or individual emergency, is down to some reduction in choice as typified by the act in question. To be clear, I am not outright rejecting a character theorist's conception of situations of extreme pressure, but Duff's hypothesis is that we do not have to.

Taking this synthesis proposal further, Williams suggests that capacity, or rather its absence, can be attributed to either a person's character, their lack of virtue, or their fear emotion.[138] Emotions are important, says Williams, because they represent a link between capacity *and* character, which are both necessary conditions for determining criminal liability. She argues that a lack of choice can indicate a reduction in (moral) capacity, sufficient to negate any imputation towards the agent's general character or virtue, on either of two understandings: that it was a limited choice constrained by existing circumstances and forced upon the actor; or on the basis that no choice existed at all because the actor was simply reacting rather than choosing to act in response to the situation.[139]

---

[135] Duff, "Choice, Character, and Criminal Liability", *op cit*. at 378-9.
[136] *Ibid* at 364.
[137] *Ibid* at 378.
[138] Williams, *op cit.* at 9.
[139] *Ibid* at 8.

In other words, and in contradiction to Brudner's argument above, this second understanding of capacity as informing questions of character and culpability through the subject of emotion appears to conform to the voluntarist denial of attributability thesis when a person becomes so overwhelmed by emotion that they cannot be said to choose (or desire) at all. This lack of choice or capacity can be influenced by emotion since, for example, fear of a threat may cause an impairment whereby the actor is less able to assess the authenticity of a threat or how best to avert it.[140] In other words, the claim is that emotions can directly impair an actor's *capacity to choose* their conduct,[141] in a way which is analogous to other denials of attributability such as mistake, involuntary intoxication and mental disorder.

We saw above, however, that even Moore was willing to admit that emotions cannot be said to incapacitate our choices,[142] and there are good reasons for thinking that a purely subjective emotional reaction should not provide the basis for determining an actor's liability. This pushes us towards the first understanding that, at best, the actor engages in a limited choice which is constrained by the circumstances and forced upon them. Williams thinks, however, that the distinction is one of little importance – both understandings show that emotions play a key role in deciding whether the actor chose to act.[143] Be that as it may, if we think that at least some actors who have their choices limited rather than vitiated by emotions deserve exculpation, we must find some other way to distinguish reduced capacity from *no* capacity.

Indeed, Duff points out that there must be something more to this concept of 'capacity' than an agent's appreciation for the empirical aspects of their actions and circumstances, as well as the ability to rationalise about what courses of action will serve whatever ends they have. This is because, as discussed above, our intuitions tell us that an actor who is seemingly stoic in the face of a situation of extreme pressure still deserves some form of exculpation on the basis that they 'could not have resisted' the threat.[144] Equally, it seems short-sighted to suggest that all that is required for exculpation is proof of an overwhelming emotion, since emotions may be triggered by abhorrent stimuli that the law should not endorse.[145] Duff thus reasons that the idea of 'capacity' must be normative,

---

[140] Duff, "Choice, Character, and Criminal Liability", at 356.
[141] See also Hart, *op cit* at p.152.
[142] Moore, *op cit.* at 559.
[143] Williams, *op cit*. at 9.
[144] Duff, "Choice, Character, and Criminal Liability", at 356.
[145] This is the main concern of Berger, who points to examples like 'homosexual panic' and racial stereotyping in instances of self-defence as being inappropriate emotional triggers for the law to recognise: see Berger, *op cit*. generally. See also C Lee, *Murder and the Reasonable Man: Passion and Fear in the*

rather than empirical, in nature.[146] Kadish alludes to a similar conclusion when he states that people suffering from a constrained choice should not be condemned because "they have simply conducted themselves in exactly the same way as most of us would in such unusual circumstances".[147]

We see this in the duress of circumstances defence where it is sometimes not enough that a defendant *actually* could not resist the threat: if a 'reasonable person' could have so resisted, then there is legally speaking no 'overborne will' for the purposes of determining liability. However, the English approach to this kind of capacity is a neutered one: although it starts from an empirical premise – liability based on an assessment of the actor's (appreciation of) certain values – it then rejects those considerations which would help to accurately assess "the relationship between the agent's actions and [their] attitudes, concerns, and values" and "whether or not [their] action displayed an improper indifference or lack of commitment to others' rights and interests, or to the law and its values".[148] As aforementioned, Williams sees this inability to appreciate the values of an actor as a consequence of rejecting the role that strong emotions such as fear should play in establishing whether and how the offender has chosen to act.[149]

Indeed, this can be seen as part of a broader attitude, as highlighted by Fletcher, whereby the courts have a general aversion to recognising defences which relate to the character of the actor rather than to the quality of the deed.[150] He argues that the common law reliance on a 'reasonable person' standard in defences such as necessity and coercion relates to this aversion, and that an emphasis on deeds often results in these defences being recognised in all cases in justificatory terms, further obfuscating their meaning when the act constitutes an all things considered wrong and the dialogue must then shift to excusing conditions. One can recharacterise this issue in the terms used here to state that courts are far more willing to recognise defences stemming from a constrained choice (i.e. the method of assessing situations of individual emergency) than they are those stemming from a constrained will (i.e. emotional reaction).

Criminal Courtroom (New York University Press, 2003) generally, and specifically at p.67ff. (on homosexual panic) and p.137ff. (on racial stereotypes).

[146] Duff, "Choice, Character, and Criminal Liability", at 358.

[147] SH Kadish, "Excusing Crime" (1987) 75 *California Law Review* 257 at 274. This broadly appears to be the conclusion Gardner reaches as well when discussing the 'gist' of excuses: *op cit*. See also Williams, *op cit*. at 4.

[148] Duff, "Choice, Character, and Criminal Liability", at 361.

[149] Williams, *op cit*. at 9.

[150] GP Fletcher, "The Individualization of Excusing Conditions", *op cit*. at 1272.

If emotions are to underpin situations of extreme pressure then, further inspection of the role of emotions in the criminal law, and specifically their role in the culpability of actors who commit *prima facie* wrongful acts, is necessary. In particular, a defence based on emotional response must be broad enough to encapsulate this normative concept of a constrained will, regardless of the multifaceted ways emotions can manifest and be presented (which are empirical concerns). We have spoken briefly above about the 'reason responsiveness' of emotions and whether it can be said that emotions play a role in reducing the capacity of an actor, but it remains to discuss exactly how emotions can interact with capacity and choice in a normative sense. Are emotions best understood as behavioural in nature, such that they invalidate capacity in a physiological/empirical sense as the voluntarist language in court judgments suggest? Or are emotions better understood as responsive to reason, and hence normative in nature? On which basis, if at all, does the Scottish criminal justice system understand emotions in substantive criminal law defences? I shall address these issues in the following two chapters.

5.4 Conclusion

In this chapter I introduced two conceptualised situations – situations of extreme pressure and situations of individual emergency – with the purpose of providing a new normative framework which would highlight the types of conduct that we think reactive defences in criminal law should encompass and why. It was proposed that these normative concepts might also be used to locate an appropriate distinction between necessity and coercion in Scots law, based on the different senses of constraint that each represents – i.e. whether the actor suffered from a constrained choice (situations of individual emergency), or a constrained will (situations of extreme pressure) respectively. These different senses of constraint are necessarily linked to the different factual nexus each situation represents: a constrained choice focuses on the circumstances the accused found themselves in; and a constrained will places emphasis on how circumstances affected the accused specifically.

The second half of this chapter then sought to explore this distinction in more detail. While the idea of a constrained choice in the context of situations of individual emergency is more intuitive, being grounded in both objectivity and the conventional understanding of choice, I questioned the current interpretation of an 'overborne will' in the current defences which sees this will as being, in many ways, also synonymous with a lack of choice. There is a distinction to be maintained between those whose choices are

limited, and those whose will is overcome in some way. While a constrained choice might neatly correspond to choice theories of culpability, it was argued that a better way to understand a constrained will which exculpates was through its emotional content; emotions tying together key themes in a more diverse culpability theory, such as choice *and* character.

This necessarily requires a robust theory of emotions which can accommodate such normative judgements. Importantly, it must be able to account for the following hard cases. First, it must be able to explain how a defence based on emotions can be available to a person who does not outwardly display an emotional reaction but nevertheless acts within the normative scope of a situation of extreme pressure. Second, it must be able to explain why a person who is clearly overwhelmed by emotion should not be granted a defence when their actions do not conform to a base level of reasonableness. In the following chapters, I shall critically examine the role of emotions in culpability for situations of extreme pressure in two ways. First, I shall dissect the occurrence of emotions in reactive defences from different jurisdictions to determine how the law currently understands them in substantive defences. I shall then analyse theories of emotion in law which are based on neuroscience to demonstrate how a defence based on emotional reaction can accommodate the hard cases mentioned above, thereby providing an appropriate basis for normative judgement in law.

6. Emotions and Reasonableness in Criminal Justice

6.1 Introduction

Chapter five outlined a normative framework for understanding reactive defences which divided an actor's potential responses into those which feature a constrained choice (situations of individual emergency), and those which feature a constrained will (situations of extreme pressure). It was argued that there is a relevant distinction between having a constrained choice and constrained will, with the latter indicating the presence of some strong emotional factor relevant to the moral blame assessment. This distinction is grounded in the voluntariness principle as understood by Wertheimer as different forms of constraint.[1] It was argued that constrained volition is distinct from non-volition, with the latter involving some constraint on capacity. Framing the dichotomy in this way allows us to restructure the defences of necessity and coercion from within the current legal framework, rather than operating purely in the realms of normativity. This is because voluntariness is an important concept in law, seen as one of the founding hallmarks of criminal liability. Indeed, moral blame can only be ascribed to voluntary actions,[2] and thus any defects with an actor's volition can render moral blame inappropriate. The presence of emotions, having an influence over our will and corresponding actions, might therefore be regarded as an important factor in determining whether an act was voluntary in the sense of being nonvolitional. Likewise, the existence of an emergency situation which forces an actor to choose between two harms provides a different volitional quality to the conduct undertaken when compared to someone who acts freely.

We can understand the blameworthiness of persons who act in situations of individual emergency by analysing the context and circumstances that they found themselves in. Culpability in situations of extreme pressure is harder to define because exculpation here stems from the accused's psyche. We must therefore consider to what extent emotions can influence a person's will – both in terms of how they affect a person and to what degree – and how this can be distilled into a workable, normative principle.

---

[1] See section 5.2.5 above.

[2] Gordon, *Criminal* Law, at paras 3.08-3.12; A Wertheimer, *Coercion* (Princeton University Press, 1987) at p.4. HLA Hart conceives of the law as a choosing system, from which punishment depends on the voluntary decision of an actor to do what the law prohibits: *Punishment and Responsibility* (1968) at pp.46-49.

The answers to these questions will entirely depend on our theory of emotions and their role in criminal law. If we regard emotions as creating a near automatous state in a person, the incapacity will be great and the rules that ought to apply should reflect this. Conversely, if emotions are capable of rational thought such that they are formed by and embody the beliefs and values generated by our appraisals and evaluations of the world around us, stricter normative requirements become appropriate.

The following two chapters therefore aim to build on the idea introduced in the last chapter of situations of extreme pressure being based on an emotional response, one which is responsive to reason and thus an appropriate site for normative judgement. This chapter serves as an introduction to the topic of emotions in criminal law defences by outlining the understanding of emotions in reactive defences currently seen in different jurisdictions, including Scotland. Specifically, this series of case studies explores how emotions have been perceived by the courts and in general folk psychology. Space precludes an holistic analysis of emotions in criminal law generally, and this thesis is only concerned with how far emotions impact a legal system's understanding and thus formulation of substantive criminal law defences. Chapter seven shall thereafter consider the equivalent understanding of emotions in defences from neuroscience before concluding on the normative basis for necessity and coercion in Scots law.

Particular attention shall be paid to the emotions of fear and anger. The former is most often associated with the defences of necessity and coercion/duress in legal systems, whereas the latter is a prominent aspect in provocation and other loss of control style partial defences. We will see that the treatment of these emotions has tended to be inconsistent across various legal systems, with anger being placed more prominently in legal blame assessments based on a particular understanding of how that emotion can be said to impact on an actor's will and capacity to commit an offence. In other words, anger tends to form the basis for a (partial) defence when it is present, whereas fear tends to be regarded as merely an indication of credibility. The legitimacy of this distinction will then be considered further in chapter seven which analyses psychological accounts of emotions in the legal context to determine if, descriptively, Scots law as it currently stands represents an accurate depiction of what contemporary science believes occurs in an actor's mind when they experience a strong emotion.

6.2 The Role of Emotions in Contemporary Legal Systems

> "An actor overwhelmed by fear cannot be expected to deliberate about his options, and consequently he cannot be blamed if he acts badly."[3]

To what extent then do emotions play a role in our evaluation of an actor's response to situations of extreme pressure? One can argue that everything we do is fuelled by emotions in some way[4] – we pursue objects and goals that make us happy and avoid things that make us fearful or sad. It is clear, however, that emotions become of particular importance in the context of reactive defences where an emphasis is placed on the psychological and physiological responses actors have to stressful stimuli as a basis for determining blameworthiness and, ultimately, legal responsibility. Even in situations of individual emergency as envisaged here emotions are relevant – we are understanding of such actors, at least in part, owing to the stress and anxiety caused by the circumstances. Thus, unpacking the concept of emotions in law and how this relates to the defences of necessity and coercion not only helps us better understand the appropriate level of blameworthiness to be applied to an actor pleading either defence, but also informs and directs the appropriate limits and requirements for each defence.

Legal understandings of emotions have historically been unprincipled and incoherent. Spain, discussing English law, notes that one of the biggest challenges for the law is explaining why it allows full or partial exculpation based on some emotions such as anger, but not others such as fear.[5] There are numerous examples of the inconsistent approaches taken by many legal systems towards their definition of the voluntariness principle and their understanding of the emotions of anger and fear in criminal law defences. The first part of this analysis highlights these inconsistencies through a focus on the folk psychological understanding of emotions in criminal law – specifically, how emotions are understood and applied in substantive criminal law defences.

To this end, this chapter will look at four case study jurisdictions as a means of developing the theme of emotional impact as a potential site for exculpation in criminal law defences: England, Scotland, South Africa and Canada. Space precludes a large-scale case study, and so English law is chosen due to its close ties to the Scottish legal system and the clear influence it has had on the Scottish approach to reactive defences generally;

---

[3] CO Finkelstein, "Duress: A Philosophical Account of the Defense in Law" (1995) 37 *Arizona Law Review* 251 at 266, explaining the classic voluntarist position towards duress.
[4] "A day will not go by without our feeling a range of emotions": A Reilly, "The Heart of the Matter: Emotion in Criminal Defences" (1997) 29 *Ottawa Law Review* 117 at 121.
[5] Spain, *op cit* at p.66.

South African law is included owing to the unique development of their law of provocation to include 'extreme emotional stress' as part of a defence which can provide a complete acquittal based on an understanding of emotion as an impulse which removes capacity; and finally Canada has been selected to provide a fuller discussion of the (in)famous 'moral involuntariness' concept[6] which has been utilised to provide one of the most generous interpretations of emotional impact in law.

Examination of these jurisdictions demonstrates that while voluntarist language is used to describe all reactive defences, there is no attempt to distinguish between different types of voluntariness, and this is despite clear evidence of preferential treatment being shown towards anger as an emotion which can impact on an actor's capacity and/or moral blame, versus the impact that fear has on mitigating guilt. The rationality of this approach is then explored in chapter seven, as well as considering to what extent normative and legal judg(e)ments are valid when aimed at persons suffering from strong emotions generally.

6.3 England

In English law, the partial defence of loss of control[7] can be regarded as incorporating emotions such as anger and fear into the rationale for leniency in murder charges.[8] In contrast, the English courts have tended to treat duress as a defence which is, on the whole, indifferent to the presence of strong emotion, despite the language used by the courts. Originally 'provocation',[9] loss of control provides a partial defence to a murder charge[10] where the accused suffers from a loss of self-control[11] which had a qualifying trigger,[12] and a person of the defendant's age, sex, and with a "normal degree of tolerance and self-restraint" in the circumstances of the defendant might have reacted in a similar way.[13] A qualifying trigger is defined in terms of the loss of self-control being attributable to either or a combination of:[14] fear of serious violence from the victim against the defendant or another identified person;[15] and/or things done or said which constituted

[6] See sections 4.4.2 & 5.2.4 above.
[7] Loss of control replaced provocation as a partial defence to murder under s56 Coroners and Justice Act 2009 (hereafter CAJA).
[8] Insofar as the criminal law can be said to recognise emotions at all.
[9] The defence was codified by s3 of the Homicide Act 1957 until repealed by CAJA, s56.
[10] s54(1) CAJA. A successful plea reduces the charge from murder to manslaughter.
[11] s54(1)(a) CAJA.
[12] s54(1)(b) CAJA.
[13] s54(1)(c) CAJA.
[14] s55(5) CAJA.
[15] s55(3) CAJA.

circumstances of an extremely grave character which would cause the defendant to have a justifiable sense of being seriously wronged.[16]

The most notable difference in the new formulation from the old defence of provocation is the removal of any reference to a 'reasonable' person standard. This standard has been replaced with guidance on what characteristics and other factors the jury may consider when assessing the defendant's behaviour. Prior to the introduction of the loss of control defence the court in *Attorney General for Jersey v Holley*[17] had already stated that assessing loss of self-control in provocation was an objective assessment, and the new prescriptive test introduced by s54(1)(c) CAJA does not appear to make considerable changes to this standard.[18] Loss of control is also narrower in scope than the old provisions, which did not limit the types of conduct which could constitute provocation (sexual infidelity is now to be disregarded in determining whether there was a qualifying trigger: see s55(6)(c) CAJA),[19] and the gravity of the conduct required to trigger a loss of self-control is now higher.[20] Generally speaking, however, both provocation and now loss of control partially exculpate on the basis of an extreme emotional disturbance, such that the ability to restrain oneself is overcome by emotional passion.[21]

The relationship between duress and loss of control, in terms of their rationale, has previously been considered in the leading authority on duress as a defence to a charge of murder, *R v Howe*.[22] Two questions of relevance arose for the House of Lords to consider regarding features of the then provocation defence and their applicability by analogy to duress: whether there was any objectivity to the emotional component of duress claims (reasonable fear versus an honest and genuine belief); and whether duress could reduce

---

[16] s55(4) CAJA.
[17] [2005] 2 AC 580.
[18] The reference to "in the circumstances of [the defendant]" in s54(1)(c) CAJA might be open to a more subjective interpretation. Consider, for example, *R v Rejmanski (Bartosz)* [2018] 1 Cr App R 18 which discussed whether mental conditions like PTSD and personality disorders should be considered relevant to the circumstances of the defendant. The court held that such conditions would be irrelevant, insofar as they reduced a defendant's general capacity for tolerance or self-constraint. This seems to mean that while such conditions may be relevant for the purposes of establishing a qualifying trigger, they may not then be used to influence the 'normal capacity for tolerance and self-constraint' requirement. Likewise, the voluntary consumption of alcohol was found not to form part of the 'circumstances' for consideration in *R v Asmelash (Dawit)* [2014] QB 103. In contrast, the court in *R v Clinton* [2013] QB 1 held that sexual infidelity could form part of the circumstances, provided that it was not the main cause for the loss of self-control.
[19] Although *R v Clinton* [2013] QB 1 holds that sexual infidelity can help establish a loss of self-control even if it cannot provide the only grounds for a qualifying trigger.
[20] Compare s3 Homicide Act 1957 with s55 CAJA, and see comments per Lord Judge CJ in *R v Clinton* [2013] QB 1 at 10, para [11] and again in *R v Dawes* [2014] 1 WLR 947 at 963, para [60]: "For the individual with normal capacity of self-restraint and tolerance, unless the circumstances are extremely grave, normal irritation, and even serious anger, do not often cross the threshold into loss of control."
[21] *R v Clinton* [2013] QB 1 at 33, para [128].
[22] [1987] AC 417.

murder to manslaughter. The court ultimately answered these questions in the affirmative and negative respectively, but there are several inconsistencies present in the speeches of the Law Lords which demonstrate a lack of normative coherence with respect to these defences.

In the first instance, the speeches of Lords Hailsham and Mackay, read alongside each other, are ambiguous on whether or not provocation is analogous to duress. When Lord Mackay considers the question of duress as a partial defence, for example, he cites authority to support the view that duress could reduce a charge of murder to manslaughter in a similar vein to provocation, although he ultimately dismisses this possibility in the instant case as being inappropriate "in the present state of the law".[23] Likewise, and towards the end of his judgment when considering the question of reasonableness and objectivity in the test for duress, he cites with approval a passage by Lane LCJ in *R v Graham (Paul)* supporting the existence of an objective criterion as a way of limiting the defence for both public policy and consistency reasons, because "provocation and duress are analogous".[24] The judgment of Lord Mackay therefore demonstrates, at several places, an openness to the provocation/duress analogy which he sees as being practically limited by how the law has developed in relation to each defence. In contrast, Lord Hailsham categorically rejected the analogy between duress and provocation in charges of murder, stating that a reduction in charge from murder to manslaughter in duress cases was a proposition he was "quite unable to accept".[25] He stated that provocation is a concession to human frailty on the basis of an emotional loss of control, whereas duress is admitted on the basis that coerced conduct is a "conscious decision" and may be "coolly undertaken".[26] In other words, they are not analogous because they have different rationales for their exculpatory effects.

The normative ambiguity created by these polarising aspects of each Lords' judgment feeds into a larger issue about the inconsistent approach taken to the emotions of anger and fear in English law. Reilly argues that Lord Hailsham's judgment serves to draw a distinction between anger under provocation as being based on a physiological mechanism, and fear under duress as being understood in cognition.[27] Consequently, there is a disparity in the level of self-control expected of an accused acting under extreme anger

[23] [1987] AC 417 at 455-456.
[24] [1982] 1 WLR 294 at 300.
[25] [1987] AC 417 at 435.
[26] *Ibid*.
[27] A Reilly, *op cit*. at 147.

and extreme fear, with the latter being expected to retain the ability to cognitively appraise the situation even in the most dire circumstances.[28] While it is no doubt correct that there is a disparity between the treatment of anger and fear in criminal law defences, this analysis only scratches the surface of what is really going on. Indeed, Reilly is quick to point out that there is a disparity, but he does not tell us what *causes* this disparity. It is only by discerning the cause that we can fully appreciate how fear is understood in English criminal law defences.

What, then, is the cause for this disparity? Returning to *Howe*, despite reaching the same conclusion to the questions posed for broadly different reasons, there is one argument which is present in both judgments: both Lords Hailsham and Mackay were reluctant to impose the 'stigma of a conviction' on someone who had acted under duress, with Lord Mackay highlighting the value a complete acquittal offers an accused.[29] Both judges considered that a person who commits an offence under duress was not deserving of punishment. Further, Lord Hailsham considered that conduct undertaken in duress represents the "lesser of two evils";[30] a concept which is intrinsically linked to theories of justification. By referring to duress in the language of lesser evils, he necessarily suggests that actions performed under duress are, all things considered, permissible or even right.[31]

On such an understanding it becomes irrelevant whether or not the coerced actor suffered from a loss of self-control; what is important is that the harm committed was lesser than the harm avoided. This seems to be at odds with how duress is traditionally understood in English law, i.e. as an excusatory defence where the act was wrong but the actor should not be blamed.[32] Indeed, and in contrast, Lord Hailsham's understanding of the loss of self-control present in provocation says nothing of the resulting conduct and its permissibility; it is solely concerned with an explanation for *why the accused committed a wrong* and is thus couched in the language of (partial) excuse.

Setting aside the potential inaccuracy of describing duress in justificatory terms, Lord Hailsham's comments nevertheless point to an important distinction which Reilly seems to neglect. Indeed, he takes it for granted that duress in English law is symbolic of fear in order to ground the argument that the discrepancy between the treatment of anger

---

[28] *Ibid*.
[29] *R v Howe* [1987] AC 417 at 435 (Lord Hailsham) and 455-456 (Lord Mackay).
[30] *Ibid* at 435.
[31] It is therefore unclear on what basis duress is supposed to be regarded as a concession to human frailty rather than the accused's conduct being, all things considered, right.
[32] See section 3.3 above.

and fear in criminal law defences exists;[33] under such an assumption he can rightly surmise that the distinction is one between affect and cognition. In reality, however, the judgment of Lord Hailsham reveals that the discrepancy exists because while anger is essential to the rationale for provocation, fear is deemed to not be essential to the rationale of duress. This is despite whatever vernacular is used to explain the defence (e.g. the 'reasonable fear of retaliation' criterion). Thus, we cannot take it for granted, as Reilly does, that duress *is* symbolic of fear[34] and, indeed, that is the pertinent question to be explored.

That the distinction between anger (provocation) and fear (duress) in murder cases is one of outright rejection rather than mere ignorance[35] of emotions as the basis for the moral blame assessment is made clear by Lord Hailsham when he states that actions committed in duress may be "coolly undertaken"[36] – the presence of fear in the accused is therefore not regarded as a necessary or sufficient condition of the plea.[37] Indeed, it does not necessarily follow that the higher standard of self-control expected in duress cases as compared to provocation cases is, as Reilly suggests, a result of courts having higher expectations of those experiencing fear – rather it seems that higher expectations are demanded because fear is only an incidental result of the situation, and not one which characterises the defence, per Lord Hailsham.

This conclusion is further bolstered by the fact that 'loss of self-control', the concept that Reilly claims represents the law's understanding of the influence of emotion on human behaviour,[38] is not a requirement for the defence of duress as envisioned by English courts. The existence of objective criteria in duress such as proportionality and immediacy requirement, explained by Reilly as mechanisms for limiting the potentially wide scope of the "imprecisely defined" concept of loss of self-control, may give the impression that loss of self-control is relevant to duress in a way that is analogous to provocation (and now the statutory defence of loss of control), but this conclusion should be resisted. The immediacy requirement fulfils a separate evidential role in duress of establishing the absence of alternative courses of action and proportionality, although

---

[33] A peculiar assumption since Reilly recognises the possibility that it is not the circumstances of fear, but the absence of a rational alternative choice to complying with the threats that ground the defence of duress: Reilly, *op cit*. at p.145.
[34] At least not against a charge for murder. Cf. JM Paley, "Compulsion: Fear and the Doctrine of Necessity" 1971 *Acta Juridica* 205 at 207-8.
[35] I am not ruling out the presence of both.
[36] *Ibid*.
[37] See also *DPP v Lynch* [1975] AC 653 at 656: "By duress is meant compulsion resulting from a threat or threats, express or implied, of grave consequences. It does not include physical compulsion or the case where a person is so terrified by threats that his mental processes are incapable of working properly and he is incapable of forming any intention whatever", confirmed by Lord Simon at 709-710.
[38] Reilly, p.134.

linked to loss of control, is in fact a more appropriate requirement in situations which assume the positive existence of an actor's cognition. Indeed, Reilly notes that the requirement for a proportionate response in provocation where there is a loss of control is a normative device which is "not grounded in the logic of human experience".[39]

The rejection of fear as forming the basis of the duress defence would seemingly condemn those coerced to kill who *do* suffer a loss of self-control because of fear to a conviction for murder.[40] In effect, an accused may have their charge of murder reduced to one of manslaughter if the killing came about as a result of overwhelming anger, but not if it came about as a result of overwhelming fear. This appears to have been a point of concern for the Law Commission, who recommended the inclusion of what became s55(3) CAJA in the new loss of control defence. This provision allows the defence where the loss of self-control was attributable to the accused's "fear of serious violence", on the basis that "[the defendant] should not be prejudiced because he or she over-reacted in fear or panic, instead of overreacting due to an angry loss of self-control".[41] This provision is, however, limited to the killing of an initial aggressor in terms similar to an excessive self-defence plea: thus if the defendant killed someone out of an overwhelming fear caused by someone *else* the defence would seemingly not apply; that is unless the initial threat could be regarded as of 'extremely grave character' such to cause the accused to have a 'justifiable sense of being seriously wronged', as per the other qualifying trigger provisions.[42]

The conclusion is thus that English law does, in very limited circumstances, recognise that an overwhelming emotion can have an influence on the moral blame assessment, sufficient to reduce criminal liability. In the substantive law (as opposed to sentencing and mitigation), these overwhelming emotional reactions are categorised as a type of loss of self-control, to be dealt with under the new statutory defence in murder cases. Despite the new categorisation of such scenarios and the inclusion of fear as an emotion capable of vitiating self-control in homicide, it appears that the restriction to murder cases means that fear continues to be treated as, on the whole, inconsequential to situations of duress (or duress of circumstances). In contrast, the exculpatory nature of duress is to be found in the difficult situation the actor finds themselves in, with the

---

[39] *Ibid*, p.135.
[40] Although he very briefly entertains the idea that diminished responsibility might offer an alternative answer for the partial defence question: *Howe* at p.435.
[41] Law Commission, *Murder, Manslaughter and Infanticide* (LC 304, 2006), para. 5.54.
[42] S55(4) CAJA. It seems unlikely that such an extension of the defence would be permitted, as the Law Commission were at pains to ensure that the new defence was kept within narrow bounds: see, *ibid*, at para. 5.61ff.

presence of fear merely serving to provide evidence that the choice the defendant faced was a hard one.

6.4 Scotland

Turning to Scots law, we see a similar phenomenon where the existence of fear is treated as incidental to a successful plea of coercion, whereas a loss of self-control caused by anger is treated as "the essence" of, and thus essential to, provocation.[43] Coercion, as defined in the leading case of *Thomson v HM Advocate*,[44] focuses on the external circumstances which ground the plea – an immediate danger of death or serious injury and an inability to resist the violence[45] – whereas provocation finds its basis in circumstances which trigger an internal reaction. As with English law, the court in *Thomson* was at pains to highlight that the "will and resolution of the accused must in fact have been overborne and overcome by the threats and the danger".[46] However, this should not be taken to represent a commitment to a subjective understanding of the emotional impact of threats on accused persons, because this is qualified by the requirement that the threats be such as would "overcome the resolution of an ordinarily constituted person of the same age and sex of the accused".[47] In reality, the focus of this test as demonstrated by the case law tends to be on establishing a suitably severe threat from which the court can infer that an ordinary person would have succumbed to the threats; rather than being concerned with any overwhelming emotional outburst on the part of the accused. To this end, fear is not 'the essence' of coercion (or necessity for that matter), it is merely used as an indicator to determine what threats are sufficient to ground a defence.

The focus on external circumstances as being constitutive of coercion, while nevertheless couching the language of the defence in the terms of an emotional response with 'overborne' or 'overcome' will, generates confusion as to the underlying rationale of the defence. That the danger must have been immediate and the will 'overborne' creates the façade that the defence is based on some kind of sudden impulse, akin to the loss of control experienced in provocation cases, but in reality what is being emphasised is the principle that no alternative course of action should have been available to the accused.[48]

---

[43] PR Ferguson & C McDiarmid, *Scots Criminal Law: A Critical Analysis* (2014) at para. 21.10.3.
[44] 1983 JC 69.
[45] *Ibid* at 77.
[46] *Ibid* at 72.
[47] *Ibid* at 75.
[48] *Ibid* at 77.

The result is that the Scottish courts become liable to the criticism that they do not understand how fear works as part of the broader human experience. This can be illustrated by *Trotter v HM Advocate*,[49] where the accused was denied the defence of coercion for smuggling drugs out of fear for his father's safety, on the basis that he had a prior opportunity to inform the authorities of the threats.[50] It was therefore irrelevant that Trotter experienced fear that, on any natural understanding of the term, would be enough to 'overcome his will'. Because it did not correspond to the normative and objective 'sudden impulse' understanding of coercion espoused in *Thomson*, one based on external circumstances rather than internal dilemma, it was insufficient to provide a defence.

We encounter a similar disconnect when considering the requirement laid down in *Thomson* that the threat be of present rather than future injury.[51] Once again, this requirement serves a normative purpose of enforcing the value that one should, if given an opportunity, seek assistance from the law before committing an offence. If the threat does not represent an immediate harm, it is likely that assistance could have been sought to prevent it, and the plea shall fail. However, and again, this requirement stands in stark contrast to the reality of how fear affects our cognitive responses in terms of our motivations and value judgements. Thus, in *Cochrane v HM Advocate*[52] a threat to blow up the accused's house for failure to comply with a coercer's demands was held to be a future injury and, as a result, insufficient to ground the defence, despite evidence heavily suggesting that the accused thought the threat was genuine, that resistance was futile, and was therefore acting on this basis.[53]

However, and in the interest of balance, there is evidence that the Scottish courts might be sympathetic to a more subjective approach to the question of emotion and its influence on overbearing a person's will where that influence was absolute. In *HM Advocate v Raiker*,[54] Lord McCluskey suggested in his direction to the jury that if a person acts out of genuine fear of life-threatening violence, such that fear is their *only* reason for acting, then that person would lack the evil intention required for a crime, i.e. they would lack the necessary *mens rea*.[55] He added, however, that if the jury considered that the accused retained sufficient control over their acts to a material degree, despite the presence

[49] 2001 SLT 296.
[50] Discussed at 4.3.1.
[51] *Thomson v HM Advocate* 1983 JC 69 at 75.
[52] 2001 SCCR 655.
[53] *Ibid* at 659.
[54] 1989 SCCR 149.
[55] *Ibid* at 154.

of strong compulsion, then the jury should find the accused guilty but with the proviso that such compulsion played a part in causing the accused to act as they did.[56] There have, however, been no subsequent cases in which coercion has been treated as a matter related to *mens rea*, with every other case treating the plea as an affirmative defence. The notion of 'coercion' referred to by Lord McCluskey seems to be something novel: a different concept from the criminal law defence of coercion as it has been outlined here and in the Scottish case law, and it is unclear whether this novel understanding of coercion would be followed by a Scottish court in the future.[57]

In general, however, and as with English law, we can see a preference for normative rules which are intended to keep the defence of coercion within tight limits. Unfortunately, the formulation of these rules in the language of emotional response has led to jarring judgments where the accused's actual fear tends to be disregarded. To be clear, a normative basis for reactive defences like coercion is essential to avoid persons evading liability for displaying unacceptable emotions, but the heavy emphasis on emotional language by Scots courts belies the assessment actually being undertaken, leading to a disconnect between what the courts say is required (an overborne will), versus what is actually required (an objective lack of reasonable alternatives).

Indeed, it might be thought that Scots law is even more restrictive than its English counterpart, owing to the strict criteria laid down by Hume and followed in *Thomson*.[58] The reader will recall from chapter two that Hume considered coercion to be "somewhat a difficult plea"[59] and argued that it would be unavailable in trials for atrocious crimes unless the situation met strict criteria.[60] These criteria have since been adopted wholesale for coercion in modern Scots law and, in part due to the view of the Scottish courts that Hume's work is somewhat sacrosanct, there have been very few modifications to this adoption. This conclusion would appear to conform to the general feeling in Scottish courts towards emotions generally, explained by Lord Justice-General Rodger in *Galbraith v HM Advocate*:

> "[T]he law makes no such allowance for failings and emotions, such as anger and jealousy, to which any normal person may well be subject from time to time. They

[56] *Ibid* at 155.
[57] The lack of development here in the past 32 years since the case was heard suggests this is unlikely.
[58] *Thomson v HM Advocate* 1983 JC 69 at 75.
[59] Hume, i, 53.
[60] *Ibid*. See section 2.3.2 above.

> do not call for the law's compassion. Rather, we must master them or face the consequences".[61]

6.5 South Africa

South Africa provides an interesting case study for the analysis of emotions in criminal legal systems, owing to a series of cases emerging in the last three decades of the twentieth century which established a subjective understanding of emotional distress in the context of the defence of provocation and general incapacity. As a result of these cases, South African law has on occasion provided a generous recognition of emotions as shaping human behaviour in terms of their effect on criminal liability. Historically, provocation provided a partial excuse to a charge of murder by reducing the charge to one of culpable homicide, like many other jurisdictions with a mandatory life sentence. Equally, persons could receive a full acquittal for lacking criminal capacity (understood as vitiating the unlawful act requirement)[62] arising out of mental illness as per s78 of the Criminal Procedure Act 51 of 1977.

However, in 1981 the court in *S v Chretien* established that incapacity may also arise out of intoxication.[63] This opened the path for other, non-pathological conditions such as provocation to affect criminal capacity. In *S v van Vuuren*,[64] the court expanded this reasoning to include persons under the influence of alcohol "and other facts such as provocation and severe mental or emotional stress".[65] By 1988 the court in *S v Laubscher*[66] had established the term 'non-pathological incapacity' for any incapacity due to factors such as intoxication, provocation and emotional stress.[67] Thus, it became possible for persons charged with murder to raise the defence of non-pathological incapacity on the basis of evidence of provocation or emotional stress experienced by the accused at the time of, or before, the killing to demonstrate a lack of criminal capacity, resulting in a full acquittal.[68]

---

[61] *Galbraith v HM Advocate* 2002 JC 1 at 19.
[62] S Hoctor, "Non-Pathological Criminal Incapacity Relating to Provocation or Emotional Stress – An Overview of Developments in South African Law" (2019) 49 *South African Journal of Psychology* 177 at 178.
[63] 1981 (1) SA 1097 (A).
[64] 1983 (1) SA 12 (A).
[65] *Ibid* at 17 G-H, per Diemont AJA.
[66] 1988 (1) SA 163.
[67] *Ibid* at 167 D-I, per Joubert JA.
[68] *S v Arnold* 1985 (3) SA 256 (C); *S v Wiid* 1990 (1) SACR 561 (A); *S v Nursingh* 1995 (2) SACR 331 (D); *S v Moses* 1996 (1) SACR 701 (C).

Broadly speaking, the test focuses on two aspects: a capacity to appreciate the difference between right and wrong (mental component); and a capacity to control oneself in accordance with that appreciation (physical component).[69] This expansion of the criminal incapacity defence thus established an understanding of emotions as being capable of negating the voluntariness of a person's conduct. In addition, Hoctor has pointed out that ordinarily we differentiate provocation from emotional stress on the basis that the former involves an immediate response to a stressor, whereas a defence based on the latter flows from a longer exposure to stressful factors which gradually diminish the accused's ability to control their actions; but in the legal context these terms were being used interchangeably.[70]

S v Wiid provides the only reported example where the defence of non-pathological criminal incapacity due to emotional stress was upheld on its facts.[71] Here the appellant shot and killed her abusive and unfaithful husband. Shortly before the incident the appellant discovered that the deceased was having another affair and confronted him. The deceased denied the accusation and beat the appellant quite severely – the injuries sustained included a broken nose and tooth, and the splintering of other teeth in the process. The deceased then threatened to kill the appellant, who retaliated by picking up her husband's revolver and shooting him with it seven times. After she was heard to ask, 'what have I done?' and police confirmed that upon arrival the appellant appeared bewildered and disorientated. Her recollection of that day was vague, and she was unable to recall the shooting itself. Goldstone AJA thought that these facts suggested that the appellant had not been acting voluntarily when she blacked out and shot her husband.[72]

The South African courts' expansion of incapacity to include factors such as emotional stress essentially created a more powerful defence which merged elements of provocation and diminished responsibility.[73] However, this movement from an objective, partial excuse rule to the subjective capacity concept was not without its critics. There were concerns, for example, that where the accused provides an unreliable account of

---

[69] R Louw, "*S v Eadie*: The End of the Road for the Defence of Provocation?" (2003) 16 *South African Journal of Criminal Justice* 200 at 201.
[70] Hoctor, *op cit*. at 179.
[71] *S v Wiid* 1990 (1) SACR 561 (A).
[72] *Ibid* at 569 C-E. This case, along with the Australian case of *Van Den Hoek v R* [1986] HCA 76, are good examples of provocation cases which are based on emotions other than anger – in *Van Den Hoek* the primary emotion referred to was fear.
[73] Diminished responsibility in South African law is intended to provide mitigation in sentencing: see s78(7) Criminal Procedure Act 51 of 1977. For proponents of merging provocation and diminished responsibility in English law, see RD Mackay & BJ Mitchell, "Provoking Diminished Responsibility: Two Pleas Merging into One?" [2003] Crim LR 745.

events, then any psychiatric or psychological evidence referenced in favour of the non-pathological incapacity defence would also lack credibility.[74] Likewise, and more generally, there was a feeling that provocation should operate to mitigate only, and that people should ordinarily be able to control their emotions – the non-pathological defence thus sent out a problematic normative message.[75] The Supreme Court of Appeal therefore took the opportunity to curtail the defence in the case of *S v Eadie*,[76] although the exact extent of this curtailment is a point of contention amongst commentators.[77]

Eadie concerned a charge of murder where the accused had beaten the deceased to death in an act of road rage. The Supreme Court of Appeal rejected the accused's attempts to plead non-pathological criminal incapacity, highlighting a distinction to be made between loss of control and loss of temper and stating that the test for loss of control is the same as that for sane automatism.[78] In this sense, the court ignored the prior judgments which had focused on an accused's ability to appreciate right from wrong, and instead focused on the capacity for self-control. It also appears to equate lack of capacity with involuntariness, a point which has been questioned by commentators.[79] This conclusion appears to severely limit the practical availability of provocation as a defence, although it remains available in theory.[80]

Burchell suggests that the judgment of Navsa JA draws an implicit distinction between instances of emotional stress which build up over time, versus those of provocation which occur suddenly; the former being more condonable than the latter.[81] This is surely the correct conclusion. However, in the court's attempts to curtail the non-pathological incapacity defence, commentators have suggested that the judgment of Navsa JA goes too far in the other direction, essentially abolishing any provocation defence (short of automatism).[82] Hoctor argues that despite the judgment in *Eadie* seemingly abolishing the non-pathological incapacity defence, subsequent case law suggests that the content and availability of the defence is unchanged with respect to diminished capacity (i.e. non

---

[74] *S v Potgieter* 1994 (1) SACR 61 (A).
[75] Hoctor, *op cit*. at 179-80.
[76] *S v Eadie* 2002 (3) SA 719 (SCA).
[77] Compare, for example, CR Snyman, "Criminal Justice in a New Society" 2003 *Acta Juridica* 1 with J Burchell, "A Provocative Response to Subjectivity in the Criminal Law" 2003 *Acta Juridica* 23. See also J Burchell, *Principles of Criminal Law* (3rd revised ed, 2006) at p.436-9.
[78] *S v Eadie* 2002 (3) SA 719 (SCA), para. [70].
[79] *Supra* fn. 77. See also Hoctor, *op cit*. at 183.
[80] R Louw, *op cit*. at 204; Hoctor, *op cit*. at 183-4.
[81] J Burchell, *op cit.* "A Provocative Response" at 29. See also on this point: *S v Moses* 1996 (1) SACR 701(C).
[82] Louw, *op cit*. at 204.

provocation) cases.[83] Irrespective of the various interpretations of *Eadie*, it is clear that the court wished to move away from the stance that pure emotion, unqualified, could allow an accused to evade liability entirely. Snyman agrees with the conclusions of the judgment in *Eadie*, if not with its reasoning, and has expressed disapproval of any kind of subjective approach, arguing that the law should embrace an objective assessment of capacity and intention, based on normativity.[84]

Beyond non-pathological incapacity/provocation, it is unclear how far emotions are recognised in South African criminal law defences. The law does not attach any significance to the distinction between necessity and coercion as understood in other legal systems; both are considered examples of necessity.[85] As necessity is understood as a defence which shows the *actus reus* to be 'not unlawful', the focus of necessity is, at least on paper, therefore one of justified conduct – the emotional state of the accused, and any intense emotion they feel in such circumstances, would appear to be irrelevant. Indeed, the fact that the threat must be real (as opposed to reasonably believed to be real by the accused) lends further force to the idea that the defence does not exculpate based on factors internal to the accused.[86]

However, Paley has highlighted discrepancies in the judgments which demonstrate that necessity might be understood differently where fear is involved.[87] In *R v Damascus* Macdonald J rejected the notion that compulsion by threats would deprive a person of the ability to form the *mens rea* required for criminal liability, but suggested that fear reducing one's power to resist an impulse to act in a certain manner should be considered a mitigating factor in sentencing.[88] Likewise, Van Wyk AJ stated to the jury in *R v Hercules* that if they found that the accused "acted under duress and was induced to do so by fear then [they should] find him not guilty".[89]

Following these cases, the position was considered again in the leading case of *S v Bailey*,[90] where the court considered a charge of murder against an accused who genuinely

---

[83] Hoctor, *op cit*. at 184-5; *S v Oosthuizen* 2018 JDR 0725 (SCA), per Navsa JA at [30].
[84] Snyman, *op cit.*, particularly at 11ff. See also, generally: CR Snyman, "Is There Such a Defence as Emotional Stress?" (1985) 102 *South African Law Journal* 240.
[85] *S v Goliath* 1972 (3) SA 1 (A) at 24D. Paley notes that the term 'compulsion' is also used in case law to refer to the doctrine of necessity: *op cit*. at 207.
[86] *R v Mahomed* 1938 AD 30 at 36; *R v Damascus* 1965 (4) SA 598 (SR) at 600. Mistaken belief is dealt with separately by applying the general concept of 'lack of culpability': see S Yeo, "Compulsion and Necessity in African Criminal Law" (2009) 53 *Journal of African Law* 90 at 95.
[87] Paley, *op cit*. at 208.
[88] *R v Damascus* 1965 (4) SA 598 (SR) at 602 & 604.
[89] *R v Hercules* 1954 (3) SA 826 (AD) at 828C.
[90] *S v Bailey* 1982 (3) SA 772 (A).

feared for their own life in circumstances where objectively they were safe. Janesen JA stated that necessity may amount to a ground of justification, i.e. a defence excluding unlawfulness, or a defence excluding fault, depending on the circumstances.[91] Burchell states that in this latter capacity, necessity would act to exclude the *mens rea* based on knowledge of unlawfulness.[92] Thus, if the accused genuinely believed that he was acting in a valid situation of necessity, he would lack *mens rea* in the form of intention which includes knowledge of unlawfulness. Granted, confirmation of mistaken justification as providing valid grounds for a defence is *not* the same as confirmation that subjective emotions such as fear are to be considered as relevant to liability in such cases, but the analogy is certainly striking, demonstrating an inclination on the part of South African courts to look beyond base intention to the motives underlying an action.

Despite the above analysis, it is unlikely that fear plays (or will play) a strong role in determining situations of necessity in South Africa. However, it is not inconceivable that fear can, and presently is, represented in such situations in an indirect way. As aforementioned, cases like *S v Wiid* demonstrate that fear may operate to ground a successful plea of non-pathological incapacity, in circumstances which look very similar to necessity or compulsion as it is known.[93] In this sense, the South African approach shares similarities with the English approach and the loss of control defence. Perhaps the solution proposed here of separate defences – one which focuses on the state of mind of the accused, and another which focuses on the quality of the action undertaken – is a more appropriate way of dealing with the issue, but the South African experience certainly implies that fear experienced by an accused in situations of extreme pressure does influence the blame calculation, and that is something to bear in mind.

6.6 Canada

Questions of emotion in Canadian jurisprudence have mostly centred around the concept of 'moral involuntariness' in necessity and duress:[94] the idea that if a person can be said to have no reasonable alternatives to breaking the law, then their action will lack the free will and agency essential to criminal culpability. At a conceptual level, this

---

[91] *Ibid* at 796A.
[92] J Burchell, *op cit.* "Principles of Criminal Law", at p.258.
[93] Indeed, the understanding of fear as affecting the *actus reus* in the sense that "it affects the voluntary nature of a person's conduct" posited by Paley in 1971 in relation to compulsion is precisely how it was understood in *S v Wiid* 1990 (1) SACR 561 (A) in relation to non-pathological incapacity: see Paley, *op cit.* at 208.
[94] See sections 4.4.2 & 5.2.4 above.

rationale might situate the necessity and duress defences alongside provocation, as an analogy can be supported between 'loss of self-control' and 'moral involuntariness' on the basis that exculpation under both rests on the fact that the act is not (fairly) attributable to the actor.[95] Nevertheless, some Canadian judges and commentators have continued to draw comparisons between duress and necessity on the one hand, and self-defence on the other.[96] In this section, I will trace the evolution of moral involuntariness from its genesis in the 1984 case of *R v Perka*,[97] to its current influence on the defences of necessity and duress and what this means for the understanding of the exculpatory effect of emotions/reactive conduct in Canadian law.

In *Morgantaler v The Queen*,[98] Dickson J argued that no system of positive law could recognise any principle which would entitle a person to violate the law because, on their view, the law conflicted with some higher social value. In *Perka*, Dickson J reaffirmed this view and added that, save for the specific identifiable situations recognised in the Criminal Code where an actor is justified in committing what would otherwise be a criminal offence, to hold that ostensibly illegal acts could be validated on the basis of their expediency "would import an undue subjectivity into the criminal law".[99] As a result, the court in *Perka* held that necessity should be understood as an excuse, based on "a realistic assessment of human weakness", where emergency situations "overwhelmingly impel disobedience".[100] Thus, the Supreme Court rejected a formulation of necessity which focused on the value of the act, instead choosing characteristics of the actor as the focal point. It is important to stress, however, that the court was at pains to disassociate necessity with rightful or permissible conduct – it was not seeking to replace this value judgment with a robust framework for subjectively determining whether an actor was blameworthy. Indeed, from there, and as aforementioned in the last chapter,[101] Dickson J developed the concept of moral involuntariness which envisages a person who, owing to extreme circumstances, is deprived of a realistic choice about whether to break the law. In *R v*

---

[95] This is, to some extent, the conclusion reached by Berger when he discusses the Supreme Court's use of the moral involuntariness concept in the case of *R v Ruzic* [2001] 1 SCR 687: see, generally, BL Berger, "Emotions and the Veil of Voluntarism: The Loss of Judgment in Canadian Criminal Defences" (2006) 51 *McGill Law Journal* 99

[96] See, e.g. *R v Hibbert* [1995] 2 SCR 973; *R v Ryan* [2013] SCC 3, particularly the opinions in the trial and appeal courts; and J MacLean, N Verrelli & L Chambers, "Battered Women Under Duress: The Supreme Court of Canada's Abandonment of Context and Purpose in *R v Ryan*" (2017) 29(1) *Canadian Journal of Women and the Law* 60, criticising the Supreme Court in *Ryan* for enforcing "a formalistic protection of the boundary between self-defence and duress": at 61.

[97] [1984] 2 SCR 232.

[98] [1976] 1 SCR 616 at 678.

[99] [1984] 2 SCR 232 at 248.

[100] *Ibid*.

[101] See section 5.2.4 above.

Hibbert,[102] the Supreme Court extended the moral involuntariness concept to duress, and in *R v Ruzic*[103] the Supreme Court declared moral involuntariness a principle of fundamental justice.[104]

In 1996, motivated by the decision in *Hibbert*, Reilly and Mikus accused the Supreme Court of failing to properly synthesise the defences of duress, necessity and self-defence (a task which the court in *Hibbert* had, by its own admission, set out to do), such that it was unclear to what extent human frailties and incapacities were both understood and utilised in assessing the validity of each defence.[105] *Perka* had rejected the idea of courts endorsing individuals' value judgements in favour of offering clemency for their supposed weaknesses. However, Reilly and Mikus suggested that the levels of objectivity differed greatly between each of the key decisions for these defences, depending on how these concepts of human frailty and incapacity had been interpreted by the instant court. In *Perka*, for example, Dickson CJ reasoned that whether an action was reasonable in the circumstances could be determined by what an objective observer with "normal human instincts" would consider was absolutely necessary.[106] Reilly and Mikus suggest that this interpretation left no room for personal circumstances to influence the reasonableness test, with incapacity being reserved for those unable to appreciate the consequences of their conduct.[107]

In contrast, Lamer CJ was more open to the impact of human frailties in *Hibbert*, arguing that while strict objective tests may be appropriate in the context of determining criminal negligence (where considerations such as deterrence would be at play), the context of criminal law defences such as duress demanded a softer touch, such that relevant human frailties might be considered appropriate to imbue the reasonable person when making an otherwise objective assessment.[108] Further still, Reilly and Mikus argue that the court in *R v Lavallee*,[109] a case concerning self-defence, there understood human

---

[102] [1995] 2 SCR 973.
[103] [2001] 1 SCR 687.
[104] This was a duress case. The Court of Appeal for British Columbia has since confirmed the status of moral involuntariness as a principle of fundamental justice in necessity cases in *R v Nwanebu* 2014 BCCA 387. Cf. *R v Foster* 2018 ONCA 53.
[105] A Reilly & R Mikus, "R v Hibbert: The Theoretical Foundations of Duress" (1996) 30 *University of British Columbia Law Review* 181 at 185ff.
[106] [1984] 2 SCR 232 at 251.
[107] Reilly & Mikus, *op cit*. at 185.
[108] [1995] 2 SCR 973 at 1022, par.[61] per Lamer CJ. In his comparison of reasonableness standards, he referred to the majority judgment in *R v Creighton* [1993] 3 SCR 3, a case concerning the relevant standard of criminal negligence for homicide. Curiously, Lamer CJ was himself a justice on that case, and had dissented from the majority precisely on the basis that the test was too objective.
[109] [1990] 1 SCR 852.

frailties to influence the broader contextual circumstances, such that the effects of an abusive relationship on the accused's beliefs were relevant to determining what was a reasonable belief in the circumstances.[110] Indeed, Lavallee had shot her husband in the back of the head after he had assaulted and threatened to kill her.

However, the authors argue that *Hibbert* interpreted *Lavallee* incorrectly, placing emphasis on separate assessments for both personal and contextual circumstances.[111] On this interpretation, the oppressive/abusive behaviour was understood to influence the reasonableness assessment in two ways. First, it would provide relevant context when determining whether there was a safe avenue of escape. Second, it would inform the relevant frailties to be considered when assessing whether fear affected the accused's perceptions.[112] Reilly and Mikus explain this difference of interpretations as that between conduct labelled 'reasonable in the circumstances' and conduct labelled 'reasonable given the person's human frailty or distorted perception'.[113] This interpretation by *Hibbert* would mark a turning point towards moral involuntariness being understood in the terms of 'distorted perceptions', rather than in the purely objective terms originally envisaged by the court in *Perka*. Indeed, Reilly and Mikus note that the possibility of losing self-control because of extreme fear is never mentioned in the defences of duress or self-defence: it is always assumed that the accused remains capable of choosing whether or not to submit to the threat.[114]

Of course, Reilly and Mikus note that if human frailty is to include the effects of extreme fear, then the conditions under which it is reasonable to become incapacitated by such fear becomes central to the objective test.[115] They note that, as a concept, extreme fear does not sit comfortably with reasonableness;[116] this is because the former seeks to explain conduct subjectively, whereas the latter tries to rationalise with some degree of objectivity. It is precisely for this reason that lawmakers and academics frequently debate the integrity of the provocation defence.

This trend appeared to continue in the following Supreme court judgment of *Ruzic* in 2001. Ruzic had been coerced by gang members in Serbia to smuggle drugs into Canada, threatened with harm to her mother if she refused. Under the statutory

---

[110] Reilly & Mikus, *op cit*. at 186.
[111] *Ibid* at 189.
[112] *Ibid* at 187.
[113] *Ibid*.
[114] *Ibid* at 188.
[115] *Ibid*.
[116] *Ibid*.

construction of duress at that time, Ruzic could not plead the defence because there was no imminent danger, and the threatener was not present when the offence was commissioned. As a result, the Court had to consider, *inter alia*, whether moral involuntariness was a principle of fundamental justice under the Canadian Charter of Rights and Freedoms, such that the immediacy and presence requirements could be considered breaches and struck down as unconstitutional. The Court ultimately approved of this approach, but not before Lebel J stressed that there was a distinction between moral *involuntariness* and moral *blameworthiness*.[117] Compelled conduct was, according to the Court, not always inherently blameless – in fact, arguably the Court held that coerced conduct is *never* free from blame, as they stated that blame was attributable on a finding of the relevant *actus reus* and *mens rea*; the "initial finding of guilt".[118]

Justice Lebel, utilising language comparable with the British courts, stated instead that moral involuntariness occurred where an actor's will was 'overborne', and thus their conduct was not freely chosen. Of importance is "autonomy in the attribution of criminal liability",[119] and thus where a person acts in a morally involuntary fashion, "his acts cannot realistically be attributed to him, as his will was constrained by some external force".[120] Berger has therefore argued that this voluntarist account represents the Court's refusal to punish conduct owing to the presence of powerful emotional motivations, on the basis of an understanding of moral involuntariness which had become suffused with a mechanistic understanding of these emotions which sees them as uncontrollable triggers.[121] In other words, the trigger in duress had become analogous to the loss of self-control trigger in provocation. If emotions were of a sufficient magnitude to overcome a person's reason, they would be exculpated on *that* basis. Berger rightly points out that this would have the necessary consequence of disguising discussions about the validity of the emotion itself, serving to potentially endorse unpalatable value judgements (such as homophobia or racism).[122]

As a result of the current construction of moral involuntariness in Canadian law, there appears to be a potential over-reliance on two particular aspects of assessing morally involuntary behaviour to achieve just results. The first is the presence of the modified objective test for determining whether there was a safe avenue of escape, which was

---

[117] [2001] 1 SCR 687, at paras [32]-[41].
[118] *Ibid* at para [41].
[119] *Ibid* at para [45].
[120] *Ibid* at para [46].
[121] Berger, *op cit.* at 107. This concept is discussed in more detail in chapter seven.
[122] *Ibid* at 109-14.

introduced in *Hibbert*[123] and cemented in *Ruzic*.[124] At present the court must ask whether a reasonable person, similarly situated, would have had their will overborne as the accused did. This obviously does very little to enquire as to the normative value of the reaction, forcing the courts to either accept these abhorrent reactions, or to deny the descriptive account and claim that a reasonable person similarly situated would not have reacted as the accused did.

The second aspect is the role of proportionality between harm inflicted and avoided in these calculations. This is an aspect which was notably missing from Reilly and Mikus' critique, and has real consequences, particularly for necessity cases where this test remains an objective one.[125] In *R v Ryan*,[126] for example, the Supreme Court took the opportunity to clarify the complex relationship between the common law and statutory versions of duress. As part of this exposition, the court stressed that proportionality under both versions acted as a safety valve in determining what counted as morally involuntary conduct.[127] Of particular note is the Court's statement that not only should the difference in nature and magnitude of the harms be considered, but also there should be "a general moral judgment regarding the accused's behaviour in the circumstances".[128] The court in *Ryan* thus aligned the proportionality test in duress more closely with the equivalent test in necessity which is strictly objective,[129] emphasising the objective, reasonableness element in the modified objective test in duress.

The circumstances of *Ryan* themselves are interesting for what they imply about the current state of moral involuntariness in Canadian law. A wife attempted to hire a hitman to kill her abusive husband, and pled duress which was sustained at trial and by the court of appeal. Demonstrating how monolithic moral involuntariness had become, the Supreme Court retreated to the classic, if somewhat arbitrary, nature of the threat division, whereby the purpose of the threat in duress was to compel the accused to commit an offence.[130] On this basis duress, and more importantly the principle of moral involuntariness, did not apply to the present case.

---

[123] [1995] 2 SCR 973 at 1022, par.[61].
[124] [2001] 1 SCR 687. In particular, see para [98].
[125] *R v Perka* [1984] 2 SCR 232 at 259; *R v Latimer* [2001] SCC 1 at para [34]. Indeed, given the affirmation of moral involuntariness as a principle of fundamental justice by the British Columbian Court of Appeal in *R v Nwanebu* [2014] BCCA 387, there may be a future challenge to this objective standard, which currently incorporates no aspects of the accused's circumstances or characteristics.
[126] [2013] SCC 3.
[127] *Ibid* at paras [70]-[74].
[128] *Ibid* at para [72].
[129] *Ibid* at para [70]ff., citing *R v Perka* [1984] 2 SCR 232
[130] *Ibid* at paras [20] & [29].

The Canadian experience is an important one, as it demonstrates how easily considerations of emotions can overwhelm the legal process if they are not firmly based in normative values, which manifest as legal requirements. The Supreme Court appears to have recognised this fact in *Ryan*, with greater emphasis being placed on proportionality and the setting out of formalistic rules, much to the chagrin of those who considered the case to be a regression from its ancestors in *Lavallee* and *Ruzic*.[131] Indeed, a balance must be struck when the law considers emotions, somewhere between their causing complete incapacitation and their pure rejection, and this can only be based on a more fuller understanding of how emotions affect us. This will be considered in the next chapter.

6.7 Conclusion

In this chapter we have taken a snapshot of various legal landscapes in the context of criminal law defences to determine how emotions are generally regarded in different jurisdictions from a folk psychology perspective. We have seen that in all cases there is a discrepancy regarding how defences like provocation which rely on anger as the source of motivation have been treated, versus other reactive defences like coercion/duress or necessity which feature fear. In more recent years there have been attempts by different legal systems to reconcile the fact that a person may be just as overcome by fear as by anger, and attempts have been made by the legislator and/or courts to accommodate for this fact. In all jurisdictions it is either stated or strongly insinuated that, where recognised, an overwhelming emotion operates to reduce the capacity of the accused. There have been attempts to temper this approach with objective, normative elements. More often than not, however, these attempts result in conceptual incoherence, where the law continues to suggest that for a plea an accused must demonstrate a lack of capacity, whereas in reality what is often sought is proof of a reasonable reaction in the circumstances irrespective of an actor's capacity.

For the most part, the conclusions reached by each legal system are more or less sound – defences like coercion and necessity *should* have normative foundations – but the theoretical methods of reaching these conclusions leaves a lot to be desired. It is submitted that most jurisdictions fail to account for emotions in law in a satisfying manner because the underlying theory requires reconsideration. The next chapter unpacks the underlying theories concerning emotions, both in terms of their impact in the commission of offences

[131] See J MacLean, N Verrelli & L Chambers, *op cit*. generally.

and any resulting liability, to evaluate how the concept of emotions should be understood in such a way as to provide a satisfying conclusion to the arguments made in chapter five that necessity and coercion should rest on separate normative foundations, depending on whether there exists a constrained choice or will.

7. Understanding the Psychology of Emotions in Criminal Law Defences

7.1 Introduction

The previous chapter provided a snapshot of various jurisdictions' approach to understanding emotions in criminal law defences. There it was argued that each legal system operates under an unsatisfactory understanding of how emotions interact with the culpability of the accused, resulting in legal rules which are unprincipled and thus confusing. As a result, I advocate for an improved understanding of the role that emotions play in reactive defences generally as a means of discovering their contours and limits. Posed as a research question, we might ask: what specific attributes of emotions are important when assessing an actor's moral blame? The answer depends on how we conceive of emotions in terms of their formation and then impact on a person.

Given that this thesis argues for an understanding of coercion in Scots law which is based on an emotional reaction to stressful stimulus (as situations of extreme pressure), the answer to this question becomes imperative. This chapter shall thus examine the different theories of understanding emotion that have developed from breakthroughs in psychology and philosophy, dating as far back as the time of Aristotle (384-322 BC).[1] These theories have been developed and adapted for legal argument by major contributions in the nineties.[2] With these theories established, I shall provide my own analysis of how emotions can be regarded as foundational to the situations of extreme pressure concept developed in chapter five, thereby bolstering the case for the emotions dichotomy introduced by Spain as an appropriate delineation for the coercion and necessity defences in Scots law.

A wholesale analysis of emotions theories would go beyond the scope of this thesis. Nevertheless, it is hoped that by tapping into the literature on emotions theories as they pertain to criminal law defences specifically, we can better understand how emotions might influence an actor and influence an assessment of moral blameworthiness. My argument therefore assumes that there are, broadly speaking, three ways to conceive of

[1] For an excellent overview, see A Reilly, "The Heart of the Matter: Emotion in Criminal Defences" (1997) 29 *Ottawa Law Review* 117 at 123-125.
[2] DM Kahan & MC Nussbaum, "Two Conceptions of Emotion in Criminal Law" (1996) 96 *Columbia Law Review* 269; Reilly, *op cit*.

emotions in criminal law: affective, cognitive and social constructionist theories.[3] Likewise, I shall assume the truth of the basis for each theory and will not seek to question the legitimacy of the underlying science or literature which gives each theory its force. I seek to evaluate these theories, as described, in relation to each other to determine which account seems more compelling.

7.2 Affective and Cognitive Theories of Emotion

Theories of emotion are traditionally split into two categories: those which seek to explain emotion as affective in nature; and those which seek to explain emotions through cognition. The result is a broad spectrum, ranging from an entirely affective understanding which sees emotions operating as a primal, instinctive force upon the actor "more or less devoid of thought or perception"[4] at one end, to an understanding which views emotions in a cognitive light, embodying "beliefs and ways of seeing, which include appraisals or evaluations of the importance or significance of objects and events"[5] at the other. Emotion theories based in affect have proven the more popular historically, and there is a wealth of both psychological and philosophical literature attempting to explain emotions in such a way.[6] This affective understanding of emotions has also tended to be the theory preferred by law courts, as we saw in chapter six. The modern literature, however, appears to favour a cognitive approach to emotions. Kahan and Nussbaum examine these two ends of the spectrum, coining the terms 'mechanistic' and 'evaluative' to categorise the affective and cognitive theories respectively.[7]

7.2.1 The mechanistic theory

The mechanistic theory operates on the notion that emotions are psychological or physiological responses – energies that compel a person to action without embodying any beliefs or ways of perceiving objects or situations in the world.[8] External objects operate to trigger these emotions within us, so that we "feel sorry because we cry, angry because we

---

[3] Two exist at opposite ends of the same spectrum (affective and cognitive), whereas the third and final theory rejects this spectrum entirely (social constructionist).
[4] Kahan & Nussbaum, *op cit*. at 277.
[5] *Ibid* at 278.
[6] On the history of the affective understanding of emotions, see Reilly, *op cit*. at 125.
[7] Kahan & Nussbaum, *op cit*. at 278ff. (mechanistic) and 285ff. (evaluative). Reilly has described the mechanistic conception as understanding emotion in "extreme affectivist terms": A Reilly, *ibid* at 130.
[8] *Ibid* at 278.

strike, afraid because we tremble".[9] To this end, there is very little distinction between emotions and other bodily functions such as hunger or tiredness; neither has any kind of cognitive content. Kahan and Nussbaum suggest that legal accounts of moral blameworthiness under both evaluative and mechanistic theories may be influenced by either voluntarist or consequentialist leanings,[10] but that the voluntarist account has strong links to the mechanistic conception of emotions.[11]

When combined with a voluntarist account of moral responsibility, the mechanistic view serves to mitigate such responsibility on the basis that strong emotions affect an actor's ability to do otherwise than engage in wrongdoing (i.e. reduce capacity). Theories of emotion which focus on affect have been popular in law historically because the idea of a 'sudden impulse' maps well to our intuitions about emotion and how it affects us. As Kahan and Nussbaum point out, the mechanistic view seems to accurately capture the connection between emotions and passivity: "Emotions feel like things that sweep over us, or sweep us away, or invade us, often without our consent or control".[12] In this same vein, the mechanistic view emphasises the idea that emotions are external to the self – forces that do something to us, with a sense of urgency, without being part of our ultimate character.[13] To some extent, in a context in which accused persons rarely testify for fear of self-incrimination, it is unsurprising at a practical level that the law's understanding of emotions would develop to be descriptive and behavioural in nature.[14]

However, Kahan and Nussbaum urge us to examine these intuitions further to discover that, in reality, the mechanistic view is ill-equipped to explain emotional outbursts despite first appearances. This is because although theories based in affect do a good job of explaining the urgent, out of body experience that emotions create, they tend to struggle to explain other aspects of emotional experience. For example, if we return to the scenarios given above – feeling sorry because we cry, angry because we strike, afraid because we tremble – we see that each raises a further question: what caused the physiological response in the first place? If I am sorry because I cry, then why am I crying? Why do I strike or tremble? I am surely crying because something or someone has caused me to cry,

---

[9] W James, "What is an Emotion?" in CG Lange & W James, *The Emotions*, *Vol. 1* (Baltimore; Williams & Wilkins Co, 1922) pp.11-30 at 13.
[10] Kahan & Nussbaum, *op cit*. at 273. For an example, see section 6.6 above.
[11] *Ibid* at 302.
[12] *Ibid* at 279.
[13] *Ibid*.
[14] W Hirstein, KL Sifferd & TK Fagan, *Responsible Brains: Neuroscience, Law, and Human Culpability* (MIT Press, 2018) at p.78.

and thus it becomes impossible to describe the emotion as an abstract 'force' without some reference to an object to which it is directed.[15]

In establishing a connection between objects and emotions, however, we necessarily must examine the relationship between the two. Why is it that certain objects cause emotional reactions? The role of this object in emotion depends on its interpretation by the person experiencing the emotion; we only feel anger towards an object if we see it a certain way, and thus another's perspective of the same object might evoke completely different feelings, like happiness. This is to say that emotions embody beliefs about their objects, which is made apparent by the fact that changes in belief usually result in changes in emotion.[16] Thus, if a child believes dogs are dangerous, she may fear them. If her family adopts one and she learns that not all dogs are dangerous, then her prior belief is likely to be replaced with one which sees dogs in a more favourable light, and she will no longer react with fear towards them. Without any kind of cognitive element present, the mechanistic view struggles to explain this phenomenon.[17]

Likewise, as a necessary consequence of treating emotions as comparable to other bodily functions, such as states of hunger or tiredness, the mechanistic view struggles to distinguish between different emotions. If all emotions are just impulses, how is one to properly distinguish between the 'impulse' of anger, versus that of fear or grief? Indeed, this is the inevitable conclusion of denying a link between emotions and any kind of cognition, such as characteristic beliefs. Kahan and Nussbaum point out that the mechanistic view must find a way to distinguish emotions without making reference to such beliefs, "for to include the thought inside the definition is to make it a part of the identity of the emotion itself".[18]

Finally, and despite offering the dominant historical understanding of how emotions should be understood in moral responsibility in the courts, the mechanistic view cannot claim to provide a descriptive account of emotions in criminal law. Spain asks why, if what matters is the intensity of the emotion and the resulting involuntary character of the actor's conduct, we only offer a (partial) defence for some emotional reactions (such as

---

[15] Kahan & Nussbaum, *op cit*. at 282. They give the example of a person experiencing grief mourning the death of a child.
[16] *Ibid*.
[17] Reilly points out that most contemporary affectivist theories concede that cognition plays a role in the generation of emotional states, with reason being overwhelmingly influenced by biological and neurological processes – without which the actor would have no real 'motive power' to act: A Reilly, *op cit*. at 127, citing K Lorenz, *On Aggression*, translation by MK Wilson (New York: Harcourt; Brace & World, Inc, 1963) at p.248.
[18] Kahan & Nussbaum, *op cit*. at 283.

anger in provocation), but not for others (fear in coercion/duress)?[19] In part, this problem appears to be a result of the differing legal bases of provocation as opposed to defences such as duress, where the former is understood in terms of subjective emotion whereas the latter is, despite referencing fear, best understood as based on the objective circumstances that occur.[20] Attempts by some jurisdictions to expand the role of defences like provocation to include other emotions[21] serve only to add confusion to an already complicated area. This is because adding further emotional triggers to a defence which is understood as being based on the distinctly mechanistic concept of 'loss of self-control' creates a dichotomy of emotions where strong emotional responses have a moral impact in some circumstances but not in others. This is theoretically incoherent.

7.2.2 The object of emotion

Despite the above analysis seeming to roundly defeat the argument for a mechanistic conception of emotions, there is one last challenge for critics to overcome. That is, we must explain why it feels intuitive to say that some emotions seemingly *do* exist without an object – I refer to these as 'manufactured emotions'. Consider persons being treated with hormonal medication, or who are suffering from a condition which causes changes in moods or hormones (such as seasonal affective disorder), or indeed even teenagers going through puberty. In each of these examples, the manufactured origin of the emotion creates the potential for the existence of any corresponding object to be obscure. Consider the example where hormonal or bodily functions cause an actor to feel angry, but they are unaware of the emotional trigger. The actor then has no idea why they are angry; they just are. In such circumstances, what is the object to which the emotion is intrinsically linked? It is no solution to point to the fact that when each actor is reminded as to why they may be feeling this sudden wave of emotion, the intensity of the anger or frustration felt often subsides and/or is replaced with feelings of relief, because until that point occurs the actor still experiences a seemingly unwilled emotion with no obvious external point of origin or belief to anchor it.

---

[19] E Spain, *The Role of Emotions in Criminal Law Defences* (Cambridge University Press, 2011) at p.66.
[20] See sections 6.3 and 6.4. Cf. Reilly, *op cit*. at 125ff.
[21] See the Australian case of *Van Den Hoek v R* [1986] HCA 76 where the provocation defence was seemingly established by fear, the inclusion of fear as a qualifying trigger in the new English loss of control defence, and the rapid expansion of provocation in South African law to include 'emotional disturbance': discussed above at section 6.5.

There are perhaps several ways to resolve this conflict. The first is to agree with a mechanistic view of emotions and argue that emotions do not depend on the appraisal of objects to exist, but not only does this seem inherently wrong, it would involve somehow explaining emotional experience without reference to beliefs or any kind of underlying value-judgements which, as aforementioned, seems an impossible task. The second solution is to talk of emotional displacement within a theory of cognition. Moore argues that most of us, most of the time, feel emotions of a certain kind and degree with regard to certain sorts of situations, and through developing an understanding of appropriate emotional responses psychiatrists can talk of 'displacing' emotions from their real and appropriate objects to nominal ones.[22] In other words, when we do not know what the object of our emotion is, we tend to 'find' another to take its place unless and until we figure out the *real* object of our emotion. Thus, we may feel angry and irritated by many objects around us until we realise that we are, in reality, experiencing these feelings as a result of our hormones, a disorder, or because we are hungry or tired.

The third solution is to argue that the metaphysical aspect of emotions is less important and, rather, emotions should be understood from a perspective whereby they are socially constructed. Reilly argues, for example, that the bio-physical and psychological causes and effects of emotion cannot be known with any meaningful degree of accuracy, and thus emotions are more usefully explained in terms of the relationship between people in their unique spatial, cultural and historical contexts.[23] Solutions two and three shall be considered in more detail below in the context of the cognitive and social constructivist theories of emotion, but for present purposes it should be clear that the first solution – that of claiming that emotions do not respond to objects or any kind of social phenomena around us beyond a mere physiological trigger – is the least preferred and therefore should be rejected if a better solution can be identified.

Affectivist theories therefore appear to offer bite with no teeth: they play on our intuitions about what it means to feel overcome with rage, but they struggle to explain *why* we specifically feel anger or any other emotion, and *how* we are supposed to distinguish between them. This is not only unhelpful but also dangerous in the legal context: by hiding the defendant's actions behind a veil of 'unwilled' or 'involuntary' conduct, it precludes

[22] MS Moore, *Placing Blame: A Theory of the Criminal Law* (Oxford University Press, 2010), pp.182-83.
[23] A Reilly, *op cit*. at 123.

any kind of critical reflection on the normative foundation of criminal law, facilitating the persistence of regressive social arrangements and values.[24]

7.2.3 The evaluative theory

Having defined and then rejected mechanistic theories of emotion in law, Kahan and Nussbaum instead promote a theory of emotions which sits at the other end of the spectrum, which they call the 'evaluative conception' of emotions. The evaluative conception, and theories based in cognition generally, state in distinction to affectivist theories that emotions *do* embody beliefs, including the appraisals and evaluations made by actors towards the importance of objects and situations around them.[25]

Broadly speaking, all cognitive theories subscribe to the view that beliefs or judgements are necessary conditions for an emotion, so that without a belief, there can be no emotion. Most versions understand this necessary condition as an internal aspect, such that beliefs form constituent parts of the emotion.[26] More controversially, some cognitive theorists suggest that beliefs are not only necessary but sufficient conditions of emotions, such that if an actor does not experience an emotion, we may have reason to doubt that the person has the beliefs previously supposed.[27] Further still, and at the very end of the cognitive spectrum, some theorists go as far to suggest that emotions and value judgements are identical; Kahan and Nussbaum give the example of grief being a judgement about the seriousness of loss in relation to a person's well-being.[28] This theory, known as the 'identity thesis', draws its force from the fact that no single psychological or physiological factor can be identified as being absolutely necessary for the presence of a given emotion; one can be sad without tears, angry without raising their voice, etc. Thus, to take the above example of grief, what must be insisted on instead is the awareness in the actor of a serious loss, resulting in an impact on their lives.[29]

One immediate issue with the identity thesis is that it risks falling victim to the same issue (albeit from a different starting point) that plagues affective conceptions of emotion, whereby they are unable to distinguish between different emotions. If every

---

[24] B Berger, "Emotions and the Veil of Voluntarism" (2006) 51 *McGill Law Journal* 99 at 103, see also sections 6.6 and 7.4.1.
[25] Kahan & Nussbaum, *op cit*. at 285; Reilly, *op cit*. at 127ff.
[26] *Ibid* at 293.
[27] *Ibid* at 293; see also MC Nussbaum, *The Therapy of Desire: Theory and Practice in Hellenistic Ethics* (Princeton University Press, 1994) at pp.371-72.
[28] *Ibid* at 295.
[29] *Ibid*.

emotion *just is* a value-judgement, then it becomes difficult to tell them apart. The serious loss example used to signify grief above could equally give rise to anger, or any other emotion, quite legitimately; at that point the belief loses all significance as a descriptive factor, and the identity theorist struggles to explain why a certain value-judgement is indicative of one emotion and not of others.[30] However, as Kahan and Nussbaum do not rely on the identity thesis for their evaluative conception of emotion, no more shall be said about it here.

Taking the moderate cognitive condition as our starting point then, what are the advantages of treating beliefs as a necessary condition for emotions, and indeed as a constituent part? Most prominently, where affectivist theories fail to account for broader understandings of emotional experiences beyond the heat of passion, being able to point to underlying beliefs and appraisals means that cognitive accounts excel at explaining how it is we come to reach the point of disturbance. By advocating for the presence of evaluations in emotional responses, cognitive accounts are also able to demonstrate why emotions appear to always be directed at an object: emotions represent manifestations of our appraisals of those objects. Moreover, such a view can explain why the value perceived in the object is of a particular sort, varying from person to person. It explains, for example, why a person may love their dog and yet others might fear it. This particular value, say Kahan and Nussbaum, is connected to the person's well-being, demonstrating some connection or role between the object and the person:[31] cognitive accounts thus view emotions as "an expression of a personal value system".[32]

How do cognitive theories like the evaluative conception transpose to the legal setting? Kahan and Nussbaum state that it is a two-step process: first, emotions contain within themselves an appraisal of the object in question; thereafter, this appraisal may itself be evaluated to determine its reasonableness.[33] In other words, emotions are responsive to reason, and thus emotional outbursts are suitable candidates for moral judgement and legal censure. Indeed, if the claim is that emotions involve evaluative thought, we naturally begin to ask what sorts of evaluations reasonable persons would make.[34] In this context Aristotle talks of actions and emotions as being subject to the 'mean', known as a course of appropriateness that can only be found by observing what the reasonable person of

---

[30] Perhaps they do not care to distinguish them, but then the theory loses any practical significance, particularly from a legal perspective.
[31] Kahan & Nussbaum, *op cit*. at 286.
[32] Reilly, *op cit*. at 128.
[33] Kahan & Nussbaum, *op cit*. at 286-87.
[34] *Ibid* at 287.

practical wisdom would do or feel in the situation.[35] Likewise, Adam Smith stated that emotions should be evaluated from the perspective of the 'impartial spectator'.[36] If the appraisal is deemed reasonable, we might say that the emotional response in question was an acceptable one.

In contrast, an emotional appraisal may be deemed to be unreasonable in one of two ways. Either it is premised on incorrect information pertaining to the scenario, as where the underlying beliefs are wrong (e.g. I am angry with you because I incorrectly believe you have stolen something from me), or it is based on an irrational or unreasonable value appraisal itself in relation to the facts or other values (e.g. an intense reaction to a trivial event).[37] It is with this second appraisal that the law is interested: it does not seek to deny that an emotional reaction (consisting of the appraisal of an object) has taken place, rather it is concerned with whether that appraisal was reasonable and therefore appropriate. The evaluative conception, then, and cognitive views generally, reject the focus on capacity that underlies affectivist theories of emotion to hold instead that actions undertaken with intense emotion are indicative of the accused's broader character, and therefore suitable candidates for normative (and therefore legal) judgement (judgment).

Here we can begin to see how a normative theory of situations of extreme pressure might deal with those hard cases of genuine but unreasonable emotional reaction: the reaction itself is not disputed, instead the underlying value appraisal it embodies is the subject of scrutiny. Likewise, an understanding of emotions which focuses on the underlying value appraisals inherent in an action can explain why persons who appear to have limited to no visible emotional response can still be evaluated under and benefit from a defence based on situations of extreme pressure. The circumstances create the need for a reaction, and that reaction can be evaluated by examining the underlying value appraisals adopted by the accused.

Not only is the evaluative conception better placed to explain the relationship between emotions and their objects, Kahan and Nussbaum argue that it can also adequately explain the 'heat of passion' element of emotions that was the strength of the mechanistic theory. You will recall that this 'heat of passion' element could be broken down into three constituent parts: passivity of the actor, the feeling that emotions constitute an external force from oneself, and a sense of urgency. On an evaluative account we feel passive in

---

[35] Aristotle, *Nicomachean Ethics*, 1105b, 29-30.
[36] Kahan & Nussbaum, *op cit*. at 288, citing Adam Smith, *The Theory of Moral Sentiments,* (1759) (DD Raphael & AL Macfie (eds), Liberty Classics, 1982), at pp.26-7, 69-70.
[37] *Ibid* at 287.

emotion because we recognise the significance of some object(s) to our well-being, whether good or bad, over which we have little control. It is thus not the *feeling* itself but the *object* with which we have a passive relationship and, because we feel strongly about these objects which we do not control, it is for this reason that emotions feel like external forces from ourselves: "emotions register transactions with a world outside of ourselves about which we care deeply".[38] On this account emotions have urgency because we are constantly engaging with evaluations based on our most important goals and values – our fear is based on a recognition of an imminent loss of such a goal or value. This explanation, say Kahan and Nussbaum, better understands urgency than even the mechanistic conception, since the factors which bring about urgency are not of an 'unthinking sort'.[39]

Finally, if the beliefs and corresponding appraisals of objects are relevant to emotions, such that those emotions can be said to represent the expression of a personal value system, it stands to reason that emotions must in some way be shaped by social factors. Cognitive theorists do not go so far as to deny the evolutionary heritage of emotions as having a biological basis, but those like Kahan and Nussbaum argue that it is "difficult to deny… that this inherited material is shaped by society".[40] If cognitive theorists are correct, then different emotions are able to be felt at different times and in different circumstances depending on the culture or society in question. It is not difficult to think of several examples: consider and contrast the general attitude towards premarital sex in the UK from before and after World War II, or equally compare the attitudes towards premarital sex between different contemporary societies now.[41] Feelings of shock and disapproval may be replaced by indifference or even happiness, given the right sort of education or environment. Beyond variance as between societies, Kahan and Nussbaum point out that social variance exists within each society itself, known as a "plurality of emotion-systems", and indeed may even be seen within an individual as they come to appreciate different social values – how they perceive certain acts or values may well change through the course of their lifetime.[42] The possibility of a moral and emotional education provides further justification for treating emotional responses as suitable

---

[38] *Ibid* at 289.
[39] *Ibid*.
[40] *Ibid* at 296. See also JT Parry, "The Virtue of Necessity: Reshaping Culpability and the Rule of Law" (1999) 36 *Houston Law Review* 397 at 399, stating that actions undertaken out of necessity can be evaluated within social structures because actors grow up in human societies shaped by norms of expected behaviour.
[41] See, e.g. M Wiederman, "Premarital Sex" in *Sex and Society* (New York; Marshall Cavendish Reference, Vol.3, 2010) pp.663-66 at p.665.
[42] Kahan & Nussbaum, *op cit*. at 296-97.

candidates for moral judgement: persons can and should shape their emotions through moral education, and may be held criminally liable when they do not.[43]

7.3 In Search of a Reasonableness Standard

It is within this recognition of the significance of social factors on the generation and shaping of beliefs, along with the acceptance of social variance, that the evaluative view encounters its first real challenge. If, as cognitive theories state, emotions embody beliefs which themselves can be appraised to determine their reasonableness and thus appropriateness, we must inevitably ask the question: reasonable by whose standards? This question asks us to consider further the idea that persons can be emotionally educated, and what impact this has for moral judgement and our resulting understanding of reasonableness standards in the law. It also must be answered if we seek to use the evaluative theory as our underlying basis for emotions in situations of extreme pressure.

7.3.1 Moral and emotional education

First, we must expand on this idea of an emotional education and its connection to moral education. The claim by evaluative theories is that because emotions are intrinsically connected to underlying beliefs, we can alter our emotions through obtaining information which would change those beliefs, and thus can find morally culpable those who express emotions which embody harmful beliefs. In other words, a reasons-responsive account of emotions involves diachronic agency, such that we assess *how* the actor came to have the responses they do.[44] In contrast, mechanistic accounts deny any close connection between moral and emotional education. Understanding emotions as impulses, the mechanistic theory encourages an emotional education which focuses on suppression and conditioning to avoid harmful emotional outbursts, as the emotions themselves cannot (really) be shaped.[45] Evaluative views therefore endorse much closer ties to moral judgement, and specifically moral education, than mechanistic views: both views are capable of altering our moral judgement about a person's actions, but evaluative views do so in a more meaningful sense. The mechanistic theory asks us to consider whether the actor could and

[43] *Ibid* at 273.
[44] Hirstein, Sifferd & Fagan, *op cit.* at p.49.
[45] Kahan & Nussbaum, *op cit*. at 298.

should have expressed their emotions; the evaluative theory asks whether the actor should have those emotions in the first place.

This distinction has practical implications for the law and its focus, including whether the law should merely concern itself with behaviour and ignore any difference between a person who effectively suppresses their hatred and a person who ceases to hate.[46] Kahan and Nussbaum provide a poignant example of why the distinction is meaningful: even those who cherish neutrality would presumably be upset to learn that teachers were merely ruling racist behaviour off-limits without then explaining the falsity of the beliefs which lead to these prejudicial views, thus allowing their students to persist in those harmful appraisals.[47] In this sense, the evaluative theory sees emotional education as a public concern, because society at large has a vested interest in persons fostering correct beliefs (and stifling those that cause harm). Emotions motivate behaviour, and so if we are interested in regulating citizens' behaviour, we should likewise be interested in their emotional lives.[48] The evaluative view therefore has probative value in terms of explaining how and why we educate members of the community to appropriately appraise objects, and further provides reason for the law to be interested not just in behaviour, but its underlying causes.

There are two important components to emotional education under an evaluative view: education by others, and education of the self. As an example of the former, we instruct and reason with children when they display behaviour that we disapprove of. Thus, if they become angry too readily over trivial matters a parent will explain that some things are not worth losing our temper over. Likewise, we teach them to experience fear when propositioned by strangers or seek to remove that fear when it stems from the colour of a person's skin.[49] Each member of society therefore undergoes a moral education which is closely linked to their emotional behaviour growing up.[50] This fact ties into the view held by many cognitive theorists that emotions are shaped by social factors. However, education by others should be seen as parasitic on, and therefore supplementary to, education of the self. Individuals with legal capacity can, and therefore should, be held responsible for their

---

[46] *Ibid* at 300.
[47] *Ibid*.
[48] *Ibid* at 297; Hirstein, Sifferd & Fagan, *op cit*. at 235: "However, if agents truly have expansive diachronic control over their choices, then to have a deterrent effect, the criminal law ought not simply to attempt to influence synchronic decisions to commit a crime (or not), but shape citizens' agency over time such that they are law-abiding."
[49] These are examples given by Kahan & Nussbaum: see *op cit.* at 299.
[50] *Ibid* at 347.

emotions (being based on prior belief formation) and, ultimately, their character.[51] This is because it is generally up to individuals themselves to form their own beliefs about objects, in accordance with good norms.[52] In this sense, inappropriate emotional responses can be regarded as culpable failings of moral perception, rather than a "lack of will power".[53]

The active contribution of the individual then, through moral education, is enough to make legitimate ethical assessments of emotion, because individuals are capable of making critical assessments and reflecting on their actions, and are thus "obliged to be good even when those around them are not".[54] It also stands to reason then that if an individual has *not* benefited from this kind of formative education, we may wish to withhold a degree of responsibility – evaluative accounts are therefore ambivalent or even sympathetic to the 'rotten social background' rationale for mitigating liability.[55]

This does not, however, say much about reasonableness. Indeed, Kahan and Nussbaum themselves argue that pointing to our fellow citizens and the objective, reasonable person does not always provide an answer since we might have good reasons for believing that the prevailing norms are unreasonable.[56] On the premise that social variance acts to create a melting pot of value judgements and norms within a given society, a response predicated on establishing that members receive education from others to supplement their own development, and therefore legitimacy as critical thinkers and as suitable candidates for evaluative judgement, is destined to fail. If my own morality and reasonableness have been, in part, taught to me by others then we must know which specific others in order to develop a workable, objective framework. There is no obvious answer, and thus we land right back at the start. Indeed, the above argument is a very good account as to why emotions are governable, but it takes us no further in determining by what standard. The most precise we can be relying on adherence to Aristotle's 'mean' is that the reasonable, virtuous man comprises of someone older, since they have the depth

---

[51] *Ibid* at 299.
[52] Consider the example given by Julia Annas of an expert electrician who knows how to solve novel problems and can articulate to a novice why a particular solution is the right one. She argues, following a similar claim by Aristotle (*Nichomachean Ethics*, 1103a30), that expertise in a character trait operates in a similar manner, such that they can apply the trait in new or difficult situations and later provide reasons as to why the acted in that way. To this end rote memorisation and rule following represents the beginning stages of acquiring a skill, but understanding what has been taught for oneself provides the necessary intelligent flexibility for expertise: J Annas, *Intelligent Virtue* (Oxford University Press, 2011) at pp.16-9.
[53] Kahan & Nussbaum, *op* cit. at 299; Hirstein, Sifferd & Fagan, *op cit.* at pp.186-7.
[54] *Ibid* at 301.
[55] See section 5.3.1 above.
[56] Kahan & Nussbaum, *op cit*. at 297.

and insight that comes from having lived and suffered. Suffice it to say, the answer 'anyone older than the accused' is not very compelling.

7.3.2 Social constructionism and emotional reasonableness

Social constructionist theories of emotion would lean into the reasonableness dilemma, pointing out that precisely because we cannot answer this question of reasonableness neatly, we should consider the possibility that emotions are socially constructed to resolve this conflict. Social constructionist accounts reject the implication that the bio-physical and psychological causes of emotion can be properly understood with any degree of certainty, and thus emotions are more usefully explained in terms of the relationship between people and their various contexts.[57] There is therefore a larger emphasis on the context within which the actor experiences the emotional outburst, and it is this context which helps us determine the reasonableness of the actor's response. Reilly, for example, when discussing the social constructionism theory of emotions, states that while classic theories based in affect or cognition attempt to explain emotion "in terms of the 'mechanisms' which cause it", social constructionists are more concerned with the symbolic meaning of the incidence of emotion – in effect, what is the emotion indicative of, given the surrounding context?[58]

Pure social constructionism might then be seen as a rejection of the character theory which underlies most cognitive accounts of emotions theory, focused rather on the "forms of symbolic interaction", with the mechanistic aspects of emotions merely serving to demonstrate patterns of behaviour which are indicative of broader symbolic structures of human associations.[59] The evaluation of emotional responses is thus a broader question about social context, the actor becoming almost entirely separated from the assessment, save that it is from their particular features and circumstances that the context in question can be known and examined. It is therefore far more relativistic in its approach, necessarily rejecting "definitive statements about emotions and the partisan nature of the debate over their origin and expression".[60]

---

[57] Reilly, *op cit*. at 123.
[58] *Ibid* at 129. To this end, Rom Harré describes people and what they do as "products of social processes": *Social Being* (Adams & Co, 1980) at p.2. An example might be the Canadian Supreme Court's approach in *R v Lavallee* [1990] 1 SCR 852, discussed above at section 6.6.
[59] Harré, *Social Being* at p.51-2.
[60] *Ibid*.

The form of social constructionism that Reilly argues for is more modest – it views emotions as the expression of an actor's values in their particular social environment.[61] It is therefore somewhat closer to the evaluative conception of emotions in terms of its link to the character theory of punishment. Thus, in the context of provocation the idea of a 'loss of self-control' is seen as a (self-)deception, aimed at trying to find some external, or uncontrollable internal cause to explain an emotion that is embarrassing and shows us in an unflattering light.[62] Social constructionism states that this deception prevents actors from taking responsibility for their anger.[63] As a result of legal systems' focus on an affective understanding of loss of self-control in provocation cases, the possibility that concepts such as gender, sexual orientation and culture may have some bearing on the context in which the anger is expressed is not considered, thus shrouding the normative values being upheld through the defence.[64] Reilly correctly points out that this is particularly problematic since some of the most common claims of the provocation defence involve scenarios where the actor's claim is based on abhorrent social values, and in which the conduct is less socially acceptable than other conduct which would amount to murder.[65]

In contrast, a social constructionism account of loss of control would seek to understand the origins of aggression to fully appreciate the wide spectrum of scenarios and emotions which may be engaged, particularly with regards to the relationships between men and women. Reilly points out that although aggression tends to manifest as blind anger which explodes immediately after a provocative act in men, this account may not accurately reflect the experience of women.[66] He points to theories which argue that aggression is utilised by women as a means of expressing emotional states, rather than as an instrumental force used to serve particular goals, as in the case of men. In effect, women's aggression might be linked to emotions other than anger, such as fear and despair. In fact, the presence of fear may suggest that some women do not respond to provocation with aggression at all, but rather with considerations of self-preservation and de-escalation.[67]

---

[61] Reilly, *op cit*. at 138.
[62] *Ibid*, citing CT Warner, "Anger and Similar Delusions" in R Harré (ed), *The Social Construction of Emotions* (Oxford; Blackwell, 1986), pp.135-66.
[63] *Ibid*. See also R Solomon, *The Passions* (University of Notre Dame Press, 1976) at p.169.
[64] Reilly, *op cit*. at 139; Berger, *op cit*. at 103. See also section 7.4.2 below in relation to Nourse's understanding of moral judgement.
[65] Reilly, *ibid*.
[66] *Ibid* at 140.
[67] Horder therefore suggests that the provocation defence may be "poorly equipped to deal with those who are driven to act as they do out of despair": *Provocation and Responsibility* (Oxford: Clarendon Press, 1992) at p.191. Although see now the new loss of control defence in England, discussed above at section 6.3.

An understanding of provocation that assumes in all cases that a person will suffer a loss of self-control from becoming enraged is thus, according to social constructionism, woefully ill-equipped to deal with the realities of human experience. In the context of coercion, Reilly points out that the situation in England and Canada is such that although fear is a central element to the availability of the defence of duress, it is not an explicit consideration. Instead, the fear experienced is something which the actor is expected to overcome, with the law assuming that actors have the ability to choose the available lawful option(s) regardless of their emotional state.[68] As an example, Reilly highlights how the current method of understanding emotional responses in English law results in judges taking the accused out of their social contexts and imposing their own context for the assessment of the reasonableness of their fear.[69] He concludes that in the English duress case of *R v Howe*,[70] Lord Hailsham's psychological assessment that the person living in normal, peaceful conditions should be able to resist threats of death, based on his Lordship's own personal experience of being a soldier during the unique circumstances of World War II, is "dubious".[71]

Thus, to return to the earlier question posed to Kahan and Nussbaum's evaluative account asking 'reasonable by whose standards', social constructionism would respond that it will vary, on a case-by-case basis, depending on the social context. What counts as a reasonable emotional reaction for a young soldier on their first tour in a war-torn country confronted with a civilian walking towards them carrying what looks like an IED will be very different from what is reasonable for an older woman who discovers her partner's infidelity at home. This theory, and ultimately its method, of dealing with emotions in criminal law is certainly compelling, but when theory is put into practice the cracks begin to show. Indeed, the social constructionism theory is so reliant on the social context of any given case that it fails to take on any formal characteristics.

In this regard, it runs contrary to normative principles such as the rule of law and maximum certainty that many jurisdictions like Scotland strive to uphold. Returning to the provocation defence, for example, if the social constructionist is to reject the 'loss of self-control' concept that underlines the rationale for the defence, how does one establish a provocation defence? Social constructionists would have legal systems examining the social context to find the answer, but what is the practical reality here? Are we really

---

[68] Reilly, *op cit*. at 147. See generally sections 6.3 and 6.6.
[69] *Ibid* at 148, citing the judgment of Lord Hailsham in *R v Howe* [1987] AC 417.
[70] *Ibid*.
[71] Reilly, *op cit*. 149.

suggesting that judges or juries should be the 'defence makers'? If juries are to decide what provocation is, then it would become incumbent on any defence solicitor worth his salt to plead provocation in every murder case, on the off chance that the jury agrees, for whatever reason, that given the context provocation as they understood it was present. It is no answer to say that there will be some requirements, as these could quite easily run contrary to the social context trying to be preserved. An immediate response or 'cooling off' period, for example, would serve to superimpose an objective contextual analysis over situations where it is inappropriate. A commitment to social constructionist theory requires an all or nothing approach; as soon as we concede some requirements into the definition, we concede that 'mechanisms' are a useful tool for determining the existence of exculpatory conditions.

7.3.3 Reconciling subjectivity in emotions with positivity in law

If one cannot then answer the 'reasonable by whose standards' question without resort to a system of criminal law defences which struggles to adhere to any kind of practical consistency, then perhaps paying mere lip service to the notion of emotions and their influence in criminal law defences is sufficient, or even desirable. Perhaps what really matters, for the existence of a reactive defence, is that strict adherence is paid to legal requirements that ensure that such defences are only being pled in pre-conceived circumstances that warrant them; the existence of an emotional outburst is, after all, just one of many extra-legal factors that can be utilised to develop workable structures for reactive defences like necessity and coercion. One very common way of establishing such a defence, for example, is to focus on whether a reasonable alternative course of action was open to the accused that would have avoided the danger. Fingarette emphasises the importance of these legal requirements in establishing the rationale for the concept of coercion in his paper on 'Victimization'.[72]

To Fingarette, the emotive language used in coercion disguises the normative truths underpinning its rationale. In this respect his view aligns with cognitive theorists like Berger, Kahan, and Nussbaum. Indeed, the desirability of clarity in the law, particularly

[72] H Fingarette, "Victimization: A Legalist Analysis of Coercion, Deception, Undue Influence, and Excusable Prison Escape" (1985) 42 *Washington and Lee Law Review* 65. Fingarette follows a tradition of theorists who believe that coercion is a consistent concept that runs through many areas of law, and thus one might look to one area in order to inform our understanding of another (e.g. coercion in contract might teach us about duress in criminal law). For another example, see A Wertheimer, *Coercion* (Princeton University Press, 1987).

with regards to the sorts of value judgements that the law makes, appears (unsurprisingly) to be a recurring theme. In contrast to these theorists, however, Fingarette views emotions as an unhelpful distraction entirely, rather than a helpful analytical tool. Reliance on emotive language such as an 'overborne will', for example, merely serves to create an unhelpful discussion about voluntariness and free will; it suggests an 'inner', psychic event that generates unacceptable factual implications.[73]

Rejecting the language of coercion which suggests *losing* one's mind, he argues that coerced conduct clearly involves *using* one's mind;[74] coercion has nothing to do with any psychological defect of 'volitional capacity'.[75] He highlights that it is rare to see expert testimony led in coercion cases to examine the psychology of the accused: if the irresistibility of desires and impulses were of the essence of the 'involuntariness' present in coercion cases, one would expect to see the use of such expert evidence.[76] He therefore challenges the idea that there is a psychological basis for coercion defences, arguing instead that the basis of exculpation in such situations is the 'apparent reasonableness' of the accused's actions, a concept which is to be understood as comprising factors such as the absence of a reasonable alternative course of action.[77]

Indeed, for Fingarette the crux of the issue in coercion is the existence of an 'unfair' choice being placed on the accused as a result of an unlawful threat.[78] He seeks to move away from any connection between 'unfairness' and 'involuntariness', pointing out that any reference to 'involuntariness' in such situations is in fact predicated on the unlawfulness of *another's* conduct and not on any psychological state in the accused: a person is said to act in an 'involuntary' manner in coercion cases only because they have been issued with an unlawful threat which constrains their choice. Any reference to 'involuntariness' should thus be dismissed; the focus should instead be placed on the presence of an unlawful threat creating "no real choice" for the accused or, in other words, the legal status of the coercer's act.[79] With voluntariness jettisoned, coercion can instead focus on the absence of a reasonable choice, or that no reasonable alternative existed.

---

[73] Fingarette, *op cit*. at 71-73.
[74] *Ibid* at 74.
[75] *Ibid* at 75.
[76] *Ibid* at 77. Although cf. *R v Lavallee* [1990] 1 SCR 852 in Canada where expert evidence was utilised to support a claim of self-defence from a victim of domestic abuse.
[77] *Ibid* at 93.
[78] *Ibid* at 79.
[79] *Ibid* at 81.

Thus, what the accused does under coercive circumstances is reasonable, by virtue of it being the *only* reasonable thing to do. The important consideration, therefore, is the reasonableness of the choice as determined by law, rather than a discussion of the accused's psychological capacity.[80] To this end, the requirement of 'fear' that is always referred to in coercion cases is to be understood as no more than an apprehension or foresight of negative, "legally relevant" consequences;[81] this is made apparent by the fact that one who analyses the coercive circumstances in a cool and calm manner would be as equally entitled to a defence as the person who is frantic and impulsive.[82]

The 'apparent reasonableness' of the accused's action, then, stems from their position as a victim in a factual nexus where they have been issued an unlawful threat backed by a demand with which compliance is the only reasonable option. The presence of an unlawful threat is essential to the victim's innocence[83] – it is what negates blameworthiness. Fingarette fleshes out this concept by analysing situations of coercion in other legal areas, developing a broader, normative theory which he calls 'Victimization'. Reliance on these areas is made, in part, because their 'wrongfulness' requirement tends to be accentuated owing to the fact that, in non-criminal circumstances of valid legal coercion, the threat need not be unlawful.[84] In situations of economic coercion, for example, he draws attention to the fact that in such cases, as with criminal coercion, there is a focus on the reasonableness of the response to a wrongful threat and demand, in large part determined by the presence of a reasonable alternative course of action.[85] Likewise, in cases of coerced confessions, while the 'reasonable alternative' test intrinsic to criminal coercion does not apply because the act of confessing is not inherently unlawful (and therefore does not require any explanation in law),[86] he argues that coercion arises in such cases because, like the threatener in criminal or economic cases, the interrogator acts wrongfully with the intent to make a confession appear reasonable in the circumstances[87] – "but for the improper influence he would have remained silent".[88]

---

[80] *Ibid* at 82.
[81] *Ibid* at 83.
[82] *Ibid* at 92.
[83] *Ibid* at 83.
[84] *Ibid* at 84.
[85] *Ibid* at 86ff, particularly at 95-96. Although cf. Wertheimer, *op cit.* at pp.23-9 who notes the inconsistency with which English courts find economic coercion in analogous cases where there is no reasonable alternative.
[86] The wrong here being any improper techniques used to obtain the confession: *ibid* at 104.
[87] *Ibid* at 98.
[88] *Bram v United States* 168 US 532 (1897) at 549.

Thus, and in general terms, Fingarette's theory of Victimization can be understood as comprising situations where a victim provides a reasonable response to a wrongful *influence*. That is to say: any wrongful and intentional manipulation of the situation which is designed to make a certain course of action appear reasonable to another to thereby induce it will generate a 'Victim'.[89] Where a Victim does what has been made to appear reasonable in a situation manipulated by the Victimizer, they should be relieved of the legal burdens normally entailed by that act.[90] Victimization then necessarily focuses on the existence of a specific factual narrative, developed into a normative standard, to determine coercion in any given case, irrespective of what area of the law is being discussed. In this sense the Victimization theory decides what constitutes coercion entirely *ex ante.* Fingarette does not outrightly reject all extra-legal/factual considerations like psychology or morality but argues that such considerations only become relevant occasionally and derive their normative force from law.[91] As a result, the concept of Victimization strives to be highly normative and legal in nature, providing clarity and precision.

Of course, the requirement for a 'wrongful and intentional manipulation' on the part of a Victimizer necessarily precludes Fingarette's theory from applying to necessity cases as traditionally understood. Indeed, Fingarette sets out his views on the coercion/necessity distinction by suggesting that the lack of (successful) necessity pleas in the case law, in comparison to duress and coercion, can be taken as evidence that courts are more willing to entertain the latter type of defence, on the basis that,

> "there [is] something systematically significant, and more forceful, about a plea of excuse for a criminal act motivated by a danger to the actor that was designedly created for that very purpose by a wrongful human threat… as contrasted with a defense based merely on the actor having faced a choice-of-evils dilemma."[92]

He therefore locates the distinction in necessity as a choice-of-evils defence "where the defendant was not being Vicitimized" and coercion/duress as a species of the Victimization defence.[93] In other words, what distinguishes Victimization from necessity is the presence of a wrongful influence. On this basis, we can assume that under Fingarette's theory of Victimization necessity applies (irrespective of whether the threat is human in nature)

---

[89] Fingarette, *op cit*. at 105.
[90] *Ibid* at 106.
[91] *Ibid* at 91.
[92] *Ibid* at 84-5, fn.54.
[93] *Ibid*. On the necessity/coercion distinction generally, see section 5.2.5.

where the accused is not a Victim (i.e. manipulated by human conduct) and is faced with a choice of evils.

7.3.4 The problems with strict formalism

Fingarette's more general point about the deceptive use of emotive language in the coercion defence is sound,[94] and it is notable that he appears to reach similar conclusions to the current thesis with respect to the necessity/coercion distinction being based not on the nature of the threat, but rather on a perceived defect in the accused's will. However, the Victimization theory nevertheless fails at key points to provide a suitable, alternative framework. First, it is not immediately clear that Fingarette's preferred understanding of the necessity/coercion distinction is possible under his Victimization theory. This is because while "[i]t is of the essence that the Victimizer aims to work his will not merely *on* the Victim but, specifically, *through the will* of the Victim",[95] Fingarette also tells us that the definition of Victimization is nevertheless broad enough to include circumstances where a Victimizer wrongfully creates a reason for action, with the intent that this reason causes the Victim to act one way, but the Victim for that same reason acts in another way.[96]

It seems disingenuous at best to argue that the Victimizer 'works their will through the will of the Victim' where the Victim does not carry out the conduct the Victimizer seeks to induce. Indeed, the excusable prison escape example Fingarette gives as explainable under the Victimization theory demonstrates this disingenuity perfectly. Fingarette suggests that the scenario where a stronger prisoner demands sexual submission from and/or threatens serious bodily injury to another prisoner, encouraging the latter to escape from prison, represents a wrongful influence sufficient for Victimization.[97] This is not an example of traditional duress, according to Fingarette, because "the defendant was not obeying the commands of others".[98] Likewise, he suggests that viewing this as a case

[94] Indeed, this point has been argued elsewhere here: see, e.g., section 6.4 above in relation to Scots law.
[95] Fingarette, *op cit*. at 105 (emphasis in original).
[96] *Ibid* at 106.
[97] *Ibid*.
[98] *Ibid* at 107.

of necessity is a "puzzle",[99] but he never explains why. Presumably, it is because he does not view prison escape in such circumstances as risking less harm.[100]

Fingarette attempts to avoid the language of traditional coercion/duress by diverting focus away from whether the Victim 'obeys the commands' of the Victimizer. Instead, and as aforementioned, he places emphasis on the Victimizer working their will through the Victim. But for this change of emphasis to be meaningful, 'working their will through the victim' must produce a requirement, and under Fingarette's description it is unclear what that is. Indeed, Fingarette admits that the hostile prisoner's threats are made with the intent that the Victim will choose to comply.[101] If the Victim does not comply, on what basis is the Victimizer's will achieved? It would make about as much sense to suggest that a snowstorm works its will through the will of the victim when it creates a risk of exposure and forces hikers to break into a nearby cabin for shelter.

If Fingarette wishes for the focus of these cases to be on the wrongfulness of the situation then, rather than the psychology of the accused, he needs a firm answer to this question. This conundrum is also fatal to a coercion/necessity distinction which is predicated on a defect of will as Fingarette understands it. If a Victim is not required to carry out the conduct desired by the Victimizer, it becomes difficult to see how any case involving a human antagonist would not be regarded as one of Victimization, leaving any non-human case to be regarded as instances of necessity. In other words, the *actual* distinction between Victimization and necessity would remain the classical distinction based on the nature of the threat, which this thesis has argued is unconvincing.[102]

Indeed, turning to the other areas of law that Fingarette relies on to develop the wrongful influence concept, we can see similar oversights. Specifically, there is a sense that he essentially picks and chooses the aspects which make sense to his theory, while disregarding the parts that do not suit the argument. To give just a few examples, in relation to economic coercion he talks at great lengths about the similarities between criminal coercion in relation to the 'no reasonable alternative' requirement, but then seems to gloss over the fact that criminal and economic coercion diverge in relation to the nature of the harm sufficient to warrant a plea – there is no discussion as to why criminal law

---

[99] *Ibid* at 107.
[100] This is based on his understanding of necessity as a choice of evils defence: *ibid* at 84-5, fn.54. Similarly, Fingarette's view could be based on necessity being viewed as a justification rather than an excuse in US law.
[101] *Ibid*.
[102] See section 5.2.5 above.

must labour under a 'serious injury or death' requirement when economic coercion does not.[103] There are good arguments to suggest that criminal coercion *should* be available for lesser harms but as this does not suit Fingarette's needs for an explanatory theory the point is quickly raised then discarded.

Further, in relation to the unique inclusion of improper promises within the meaning of coerced confessions, he argues that a distinction is justified here since criminal acts are inherently unreasonable and the only thing that can reverse this state is the threat of a graver crime being committed if the victim should refuse to comply.[104] Not only does the attempt to rationalise criminal coercion in terms of lesser evils reasoning appear to be at odds with the premise of the Victimization theory (which is predicated on an *apparent* and not *actual* reasonableness), but equally so does the explanation as to why improper promises are rightly rejected from the scope of criminal coercion: if the underlying rationale of the Victimization theory is the presence of a Victimizer who makes a course of action appear reasonable to the Victim, then an explanation for the distinction predicated on the fact that improper promises are incapable of making a situation appear reasonable to a Victim is unconvincing. Indeed, what if the threatener and the Victimizer are separate people? If X threatens Y and then Z offers to tell X to back off if Y commits a crime of Z's choosing, should this not count as coercion? It may well not count under some current legal systems, but it is not obvious that this is a conclusion we should be rushing to enshrine in a normative theory.

Finally, and this is perhaps the most important blow to Victimization in terms of its place within a broader argument about the role of emotions in criminal law: it seems to wilfully ignore the role of emotions in an attempt to sterilise the law as much as possible, but it is unclear that this is, ultimately, desirable or even possible. Emotions are an anathema to Fingarette, and thus they are completely excluded from the discussion. The problem with such an extreme approach is that it forces him to reach insincere conclusions about why some instances of coercion are admitted. To return to the coerced confessions context, at several places he states that it would be "quite the wrong language" to characterise an accused's will as being 'broken' or their capacity 'drained' in circumstances which include repeated questioning, for days on end and even in situations where the accused was dipping in and out of consciousness.[105] Indeed, Fingarette suggests

---

[103] Fingarette, *op cit*. at 95.
[104] *Ibid* at 102.
[105] *Ibid* at 99-100 and 102-103, discussing *Culombe v Connecticut* 367 US 568 (1961) and *Mincey v Arizona* 437 US 385 (1978) respectively.

in each of these cases that the accused had a reasonable alternative of silence – this seems to completely neglect the fact that in each case this *was* the choice being actively chosen by each individual until they could no longer take being questioned anymore; silence was, in the circumstances, no longer a reasonable choice.[106]

Likewise, his argument in relation to the economic coercion of individuals (as opposed to entities) also betrays his goals. He discusses the case of *Leeper v Beltrami*[107] as an example where the court found that there was no 'reasonable alternative' to Leeper making a wrongfully demanded payment, where failure to do so would result in the foreclosure and sale of her home. While recognising the court's decision that "the unique personal value of one's own home" represented a legally relevant interest which could "not reasonably be translated into future money damages in the context of resisting wrongful demands",[108] Fingarette fails to engage with the implications of what 'unique personal value' meant in this context. Arguably, 'unique personal value' derives from sentimentality – i.e. a strong emotional connection with an object which provides a value intrinsic to that person as opposed to the outside world. Thus, far from establish that coercion is devoid of emotions, situations of individual economic coercion might provide perfect examples of where emotions are prominent.

A wholesale rejection of any role for emotions in criminal law, then, as proposed by the Victimization theory appears undesirable and unconvincing. This in turn leads us to question whether it is even possible in legalist theories generally. Such theories propose to analyse the reasonableness of the choice/response in legal terms, rather than based on the psychological state of the accused. To this end, proponents like Fingarette must go to great lengths to outline a normative theory which is based on objective circumstances that occur external to the accused. For Fingarette, this requires the presence of a Victimizer and the lack of a reasonable alternative to create a state of 'apparent reasonableness' from which the legal system can assess whether the accused's response should attract condemnation. The goal is for the test to be as neat as possible, providing absolutely certainty in every case. But it is taken for granted that we know what the exhaustive list of reasonable alternatives is, and how actors come to those conclusions. Indeed, this is a pitfall that all strictly legal theories, like Victimization, must traverse, and it is unclear that they can

---

[106] There are circumstances in *Culombe* which suggest that other factors had an impact on Culombe's decision to confess, namely his family pleading with him to do so, but Fingarette makes his argument separate from these factors, choosing to focus on the more general point that a person who is interrogated for days does not have his will 'broken' or 'drained'.
[107] 53 Cal. 2d 195, 347 P.2d 12, 1 Cal. Rptr. 12 (1959).
[108] Fingarette, *op cit*. at 89.

successfully do so, at least not without some concession to the subjective element inherent in defences which attempt to account for deviations from the standard commission of a crime.

The purely legalist account thus inevitably brings us back to the initial question, 'reasonable by whose standards', but instead of determining what contextual considerations are reasonable, we are now tasked with determining what factual considerations should populate an assessment of an accused's alternative courses of action. Fingarette is technically correct when he states that this is a different kind of psychological assessment in that it does not ask whether a victim's will was overborne or destroyed,[109] but it still engages in a degree of subjectivity which invites reasonable questions about how actors came to the conclusion that they 'had no alternatives', and the Victimization theory on its own cannot readily exclude psychological considerations from that assessment. Put generally, and following the argument of this chapter, legalist theories can move the goalposts as much as they like but must inevitably consider the emotional life of the actor and hence engage in moral judgement, it being an integral part of the criminal justice process, at some point during this legal assessment.

7.4 Determining Reasonable Emotions in Reactive Defences

So once again we return to this seemingly intractable question: 'reasonable by whose standards?' I have sought to outline the primary responses to this question above to fully appreciate the alternative views, and why it seems unlikely that they offer a suitable explanation or understanding of how we might go about answering this conundrum. The reader will recall that this question was first posed to the evaluative theory of emotions as a challenge to be overcome if we seek to utilise the theory to explain situations of extreme pressure – we have examined solutions that either involved leaning in or backing out entirely from an understanding of criminal law which places an emphasis on the role of emotions on the human experience. I have demonstrated that neither offers a suitable solution – emotions are important such that they should have a place in the blame calculation; but equally their subjective nature makes them an unfit gatekeeper to liability.

As is common with difficult questions, a degree of compromise is necessary here to determine the role of emotions in developing a reasonableness standard: the actor's

[109] Fingarette, *op cit*. at 98-99 in the context of coerced confessions.

emotional response should be reasonable by the standards of a legal system which has regard, but is not bound, to the prevailing norms of that society. This view is consistent with Berger's petition for a cognitive understanding of emotions in criminal law defences which embraces the normative judgements that legal systems regularly make.[110]

7.4.1 Normative judgement versus political neutrality

Berger's argument, in the context of Canadian law, advocates honesty in the criminal justice system when it comes to its latent normative content; an honesty which he argues is lacking in part due to the prevalence of a particular aspect of liberal theory which guides legal defences.[111] Liberal theory, as understood in the classic Rawlsian account, expresses a strong resistance to public moralising on the basis that the public sphere should be morally neutral to allow persons to pursue their own understanding of the good life.[112] On this foundation, the law can achieve an "overlapping consensus"[113] by appealing to an understanding of justice based on a shared reason which is independent of opposing philosophical and religious doctrines.[114] Berger suggests that this desire for neutrality encourages the law to adopt a mechanistic understanding of defences such as provocation and coercion, because doing so avoids engaging in the kinds of value judgements which are antithetical to a legal system committed to equality, reason, and the development of a public space in which "everyone can find some normative resonance".[115]

Given the challenges posed by multiculturalism in a society with competing understandings about the acceptability of various types of conduct, liberalism encourages an aversion to public moralising for fear that the resulting evaluative chain reaction among competing normative views might lead to a "social meltdown".[116] The liberal desire for consensus helps to create an environment that inevitably focuses on explanatory idioms like 'voluntariness' in emotion, couching the debate in terms of whether an actor reached a certain point at which capacity was compromised. This focus on capacity serves to

---

[110] B Berger, *op cit.*
[111] Berger does not suggest a conscious adoption of liberal philosophy from judges, but rather that the "latent theoretical apparatus exerts a strong centripetal pull towards mechanistic idioms like moral involuntariness": *ibid* at 120.
[112] *Ibid*, citing J Rawls, *Political Liberalism* (New York: Columbia University Press, 1996) at p.9.
[113] Rawls, *ibid* at p.11.
[114] *Ibid* at p.9.
[115] *Ibid*. See also DM Kahan, "The Secret Ambition of Deterrence" (1999) 113 *Harvard Law Review* 413 which argues that the deterrence justification for criminalisation is used as a cover to avoid any normative or moral debate on contentious topics.
[116] *Ibid* at 121.

"refrigerate the whole issue… instead of flaming the passions of the populace on issues such as domestic violence, hierarchies of sexuality, and the legitimacy of harming others to protect one's own".[117] In other words, mechanistic conceptions of defences seek to explain how and why an accused became so provoked, but shy away from considering whether it was reasonable for the accused to have done so.

Although a legal system designed to minimise social dissensus is both logical and sensible, by now it should be clear that explanatory methods of judgment based on a mechanistic understanding of emotions should be avoided because they betray the latent normative content of the law. Indeed, this desire to avoid moralisation in fact results in a deflection of moral judgement from the accused towards circumstances and/or victims. For example, it has been lamented that the law's penchant for explaining rather than judging behaviour has resulted in the legitimisation of unorthodox defences, known collectively as 'abuse excuses', which share a common goal of deflecting responsibility from the person who committed an offence onto the person, condition or circumstances who may have abused or otherwise caused the accused to act as they did.[118] By undermining the victim, who indeed may be unable to defend themselves, the accused is able to elicit the necessary sympathy through which a jury can determine that their physiological reactions were reasonable. A focus on concepts such as 'loss of self-control' helps facilitate such defences because there are numerous ways in which one can legitimately lose self-control. As Wilson states:

> "When a jury *judges* a defendant, it considers his or her mental state only to the extent necessary to establish the existence of one or another of a small list of excusing or justifying defenses, such as insanity, necessity, or self-defense. But

---

[117] *Ibid* at 122.

[118] A Dershowitz, *The Abuse Excuse: And Other Cop-Outs, Sob Stories, and Evasions of Responsibility* (Boston, MA: Little Brown & Co., 1994) at p.19. He notes that the likelihood of success for these defences increases if the finger of abuse can be pointed to the specific person whom the defendant killed or injured. See also JQ Wilson, *Moral Judgment: Does the Abuse Excuse Threaten our Legal System?* (New York: Harper Collins, 1997). Wilson's examples include the 'Twinkie' defence (the infamous case concerning the killings of Harvey Milk and George Moscone – the killer Dan White claimed he suffered diminished capacity as a result of his depression, and his consumption of Twinkies was produced as symptomatic evidence of this), the 'mob frenzy' defence (a case where two rioters claimed that mob frenzy prevented them from forming the specific intent necessary to sustain conviction for attempted murder or aggravated mayhem), and the 'baby-blues' defence (a case where a mother experiencing post-partum depression killed her child and threw her in a river before reporting her kidnapped to the police. After eventually confessing to the killing, prosecution accepted a plea of voluntary manslaughter). See also, generally, C Lee, *Murder and the Reasonable Man: Passion and Fear in the Criminal Courtroom* (New York University Press, 2003).

> when a jury *explains* the defendant's actions, it searches for a full account of the factors – the motives, circumstances, and beliefs – that caused them."[119]

Commentators like Spain suggest that by recognising reactive defences through the rationale of loss of self-control, rather than questioning the quality of the accused's conduct, the law excuses behaviour without engaging in any real judgement and thus enables the creation of 'abuse excuses' which provide unmeritorious grounds for defences.[120] The unsavoury reality of this 'sterilisation' is that successful defences on these terms nevertheless provide implied approval for normative judgements which favour the social status of certain groups within society above others. Indeed, Berger points out that even when the law tries to speak in sterilised, mechanistic terms, if it then allows provocation resulting from 'homosexual panic' to mitigate punishment, for example, it inevitably sends a strong message to certain minority groups that they are not valued in the same way as the majority.[121]

Thus, rather than sending no message at all, sterilisation actually operates to serve the one segment of the community who benefit from the status quo by creating a standard of 'normal'.[122] Through a refusal to engage in active moralising, the legal system risks condoning attitudes which demonstrate hatred towards other, often vulnerable, members of society, provided these attitudes manifest themselves in the appropriate way (i.e. as an emotional outburst resulting in reduced capacity).[123] One way that the law ends up evaluating actions in such a way, and therefore offering a particular kind of judgment, is by imposing an objective test in reactive defences so that the 'reasonable person' plays an important role in judging behaviour, even in the context of loss of self-control.[124] This reasonable person is, more often than not, a white, heterosexual male representing the "majority culture".[125] Those with such privilege are therefore able to rely on dominant social norms of masculinity, race, and sexual orientation to bolster their claims of

---

[119] Wilson, *op cit.* at p.90 (emphasis in original).
[120] Spain, *op cit*. at 17.
[121] Berger, *op cit*. at 124.
[122] *Ibid* at 126. See also V Nourse, "Passion's Progress: Modern Law Reform and the Provocation Defense" (1997) 106(5) *Yale Law Journal* 1331 at 1385: "In a world in which social norms are changing, not taking a position becomes a position, one that endorses the status quo even as it denies that it is endorsing anything at all"; Lee, *op cit.* at p.250.
[123] This can lead to problematic responses even within minority communities – see, e.g. the cultural justification for provocation or temporary insanity presented in cases of Chinese men who kill their wives: Lee, *op cit*. at p.96ff.
[124] Spain, *op cit*. at 18; for further reading on the 'reasonable man' as providing normative judgements, see JD Armour, "*Race Ipsa Loquitur:* Of Reasonable Racists, Intelligent Bayesians, and Involuntary Negrophobes" (1994) 46 *Stanford Law Review* 781, in particular at 787-90 discussing the idea that reasonable =/= typical beliefs. See also Lee, *op cit*. at pp.237-238 on equating typicality with reasonableness.
[125] Lee, *op cit.* at p.277.

reasonableness while others cannot.[126] We must therefore accept that "public moralizing is going on whether hidden behind the veil or not".[127]

Berger suggests that the focus on a singular aspect of liberal theory (moral neutrality in the public sphere) comes at the cost of a broader understanding of what liberalism should encapsulate. He notes that part of Rawls' vision of political liberalism is the centrality of the notion of a deliberative democracy.[128] This kind of democracy demands that public institutions reflect a crucial yet thin layer of fundamental political values, including equality, and citizens in turn should reflect on how far these institutions embody these values.[129] With this broader understanding in mind, Berger argues that the overzealous pursuit of consensus and neutrality by means of a mechanistic approach to emotions undermines the equally essential deliberative aspect of the liberal polity by hiding the normative aspects from view.[130] In other words, liberalism is not necessarily a bad philosophy for reactive defences, but the legal system's piecemeal adoption of parts of it certainly is.[131] In fact, our understanding of neutrality itself within the liberal idea has been a site for disagreement,[132] lending further evidence that blind acceptance of its value should be questioned.

Moral judgement is therefore an integral part of criminal law, and this is so irrespective of any institutional designs to veil it – indeed, at the end of the day veiling is all the criminal law can do since it has no power to eliminate its expressive function, which serves to uphold normative values entrenched in the system. Likewise, it is unclear that the avoidance of moral judgement is even supported on a broader interpretation of the liberal theory, which has otherwise been utilised to encourage adherence to mechanical understandings of emotions in the (misguided) pursuit of achieving true neutrality.

---

[126] *Ibid*. See also, generally, V Nourse, "Passion's Progress", *op cit*. arguing that there is a 'hidden normativity' of provocation making it harder for heterosexual women and gay men who kill their partner in response to infidelity to convince legal decision makers that they acted reasonably.
[127] Berger, *op cit*. at 124. See also CL Carr, "Duress and Criminal Responsibility" (1991) 10(2) *Law and Philosophy* 161 at 182; Nourse, "The New Normativity: The Abuse Excuse and the Resurgence of Judgment in the Criminal Law" (1998) 50(4) *Stanford Law Review* 1435 at 1461. Cf. Wilson, *supra* fn. 119 who argues that judgement has been lost, rather than merely hidden.
[128] Berger, *op cit.* at 122.
[129] *Ibid* at 123.
[130] *Ibid*.
[131] On this point see V Nourse, "The New Normativity", *op cit.* at 1457.
[132] Berger, *op cit*. at 123, citing the views of Charles Taylor that liberalism provides only the political expression of one range of cultures, being incompatible with other ranges.

7.4.2 What is a reasonable judgement?

Once we are willing to admit that the law can and regularly does make normative judgements about people's behaviour, the question of reasonableness becomes much more palatable and we can begin to unpack this concept. What does it mean for reasonableness to be determined by a legal system which in turn takes its guidance from prevailing societal norms, and how do we avoid the pitfalls identified above? Reasonableness, understood as the objective prong in a legal judgment, is often personified as the 'reasonable man' or 'reasonable person'. As aforementioned above, this reasonable person tends to be someone who closely represents the majority culture. Our current conception of reasonableness in reactive defences is therefore in need of critical evaluation. Nourse, undertaking such an endeavour, considers three ways of conceptualising evaluative judgement to fully realise the ideal reasonableness standard: 'judgment as community'; 'judgment as character'; and 'judgment as critique'.[133]

'Judgment as community' involves judging individuals not just in terms of their capacity for choice, but also for the quality of that choice.[134] Juries must therefore judge not only the ability of the accused to choose, but also on the quality of that choice. Such judgement posits that there is a 'good' to strive towards, rather than a neutrality to be achieved among different visions of 'good'.[135] In this sense, normativity has been lost and must be rediscovered, and the liberal ideal of neutrality rejected in favour of the republican theory of "inculpation", whereby a person's judgements are measured by the practical wisdom of the community, represented by juries.[136] Kahan and Nussbaum's evaluative conception of emotions is receptive to this kind of judgement.[137] It also adheres to Berger's theory which advocates a move away from the conventional understanding of the liberal philosophy. In essence, it is about judging the actor's reaction in line with shared community values to determine if it reaches the standard of 'good conduct'. This can only be done by assessing the quality of conduct because it is in this way that we can in turn assess an individual's social and political engagement (and their alignment with community goals).

---

[133] Nourse, 'The New Normativity', at 1458-67.
[134] *Ibid* at 1458.
[135] *Ibid*. Nourse relies on Huigens understanding of republican theory which supposes that there is an end/purpose served by political association, which can be contrasted with liberalism's approach which takes 'good' to be an aggregate of individual preferences: K Huigens, "Virtue and Inculpation" (1995) 108 *Harvard Law Review* 1423 at 1457.
[136] Nourse, "The New Normativity", *op cit*. at 1458.
[137] See above at section 7.3.1.

Judgment as community can be understood in two distinct ways. As aforementioned, scholars like Kahan and Nussbaum ask us to uncover the evaluative judgements which are hidden behind behaviourist trappings, whenever they may appear.[138] In contrast, Wilson understands community judgement as discoverable through a small selection of traditional excuses and nowhere else: judgement inheres in particular excuses.[139] In this respect his theory can be seen as backwards reaching, looking exclusively to conventional definitions of defences such as insanity, necessity and self-defence to provide the sole source of the law's normative content. New defences, according to Wilson, are incapable of doing this – they merely explain behaviour rather than create normative standards.[140] For context, Wilson makes these claims because he fears that the proliferation of 'abuse excuses' can be traced to the lack of normativity in legal judgment. He therefore seeks to limit the number of new excuses supposedly created by the lack of a normative gatekeeper.

However, under such a definition society must be viewed as static, incapable of any true evolution: normativity is 'complete'. Given the current understanding and requirements of the conventional defences Wilson champions, such a theory appears to support the status quo of, e.g., defending men who kill their unfaithful wives while condemning battered women who kill their abusive partners.[141] For this reason Nourse describes this kind of normative evaluation as, in reality, "judgment as history".[142] For Wilson, if one accepts the claims of battered women, one must also accept claims of racial prejudice and homophobia,[143] but this misses the point of evaluative judgment as an intellectual approach toward all excuses – one does not get to pick and choose behaviouristic explanations in some cases and evaluative explanations in others. 'Judgment as community' then, is about understanding in every situation whether the actor displayed good practical reasoning in the circumstances, not just in those select instances recognised by law. It therefore rejects the 'abuse excuse' as being a problem of invention and rather sees the issue as one of interpretation.[144]

The second conception of judgement in criminal law outlined by Nourse is 'judgment as character', an understanding which focuses heavily on reclamation of the

---

[138] Nourse, 'The New Normativity', *op cit*. at 1461.
[139] Nourse, 'The New Normativity', at 1460-61.
[140] Judgement is therefore not just hidden but has been lost. See above at fn.119, Wilson, *op cit.*
[141] Nourse, 'The New Normativity', at 1449-1456.
[142] *Ibid*, at 1461.
[143] Wilson, *op cit*. pp.59-63.
[144] With an evaluative interpretation of such defences legitimising some excuses (battered women killing their abusers) while rejecting others (killing in 'homosexual panic').

traditional concept of the 'reasonable man' in an Aristotelian sense: the man of good character rooted in practical wisdom. The theory would therefore turn on reclaiming the concept of the reasonable person from the clutches of subjectivism, where it has become so personalised that "juries are tempted to judge defendants by their self-description rather than by a common standard".[145] Nourse, however, is highly sceptical of the utility of a reasonable person standard in establishing 'judgment as character' in both the traditional and evaluative conceptions of criminal law. This is because the concept of a 'reasonable person' is normatively ambiguous, capable of encompassing the 'good man' – i.e. representing what we ought to do – but also the 'typical man' – what people actually do.[146] We saw this above in relation to the issue of determining whose evaluative standards the accused is to adhere to in the context of emotions.[147]

Further, Nourse argues that the adoption of this descriptive, reasonable person standard undermines the evaluative ideal. Giving provocation as an example, she states that over time the rationale for the defence has shifted from one which was normative (protecting one's honour based on wrongs defined by social relationships), to one based on a physiological reaction (a natural force within the blood which took time to cool), to one based on mental distress as a state of mind (focused on how a reasonable person would react).[148] The Model Penal Code's position of asking juries to identify the actor's personality characteristics essentially represents the culmination of this evolution.[149] Nourse argues that this development represents the "personification" of the defence, shifting the law's focus from an evaluation of the norms governing the conduct of the persons and relationships involved, to an analysis inside the emotional life of one person.[150] This has important ramifications for the kinds of questions we ask: historically the law was concerned with the norms governing relationships which were susceptible to critique; now the law seeks only to determine whether the accused suffered from the kind of distress which would overcome an ordinary, reasonable person.[151] Of course, such an approach implies that, in some cases, persons can and *will* lose control in the necessary sense, thus legitimising at least some instances of questionable provocation by naturalising within the minds of jurors the "answers to the very questions we are struggling to

[145] Nourse, 'The New Normativity', at 1462.
[146] *Ibid*.
[147] See section 7.3 in particular.
[148] Nourse, "Passion's Progress", *op cit*. at 1384.
[149] *Ibid*; see MPC, §210.3.
[150] *Ibid* at 1385.
[151] *Ibid* at 1387.

answer".[152] It is an "aspiration to descriptive neutrality", which results in normative challenges being rejected out of hand.[153]

Indeed, it is the difference between an evaluative statement which states that intimate relationships are so important to an individual's sense of self-worth that we should deem less culpable those who kill within such an intimate relationship (a statement which can itself be challenged for the normative values it seeks to uphold), and a statement which says that persons with particular characteristics are likely to lose self-control in such situations (which, through a focus on capacity, offers very little by way of normativity and thus challenge).[154] The result of a historically normative defence being sterilised, couched in terms of identity and a comparison with the reasonable person, has led to the enshrinement of outdated understandings of relationships. In asking whether a reasonable person would have lost control in such a way, the law becomes bogged down in questions about what characteristics the reasonable man has, essentially stymying normative debate and freezing the law's evaluative understanding of questions regarding relationships and anger (for example) to one fixed moment in time.

Lee has thus argued that if reasonableness standards are to be used at all, the law ought to recognise both positivist and normative reasonableness.[155] Reasonableness under a positivist approach is equated with typicality: it asks what most individuals would feel, think, or do if placed in the actor's situation.[156] As we saw in the previous chapter, many legal systems adopt this understanding of reasonableness. Lee argues that this conflates act reasonableness with emotion reasonableness.[157] The emotions expressed may be reasonable responses to the offending stimuli, but this does not necessarily mean that the resulting actions are reasonable. Thus, Lee argues that not only should the emotions experienced be reasonable in terms of typicality, but the resulting act should also be reasonable in terms of underlying normative values that the law seeks to express.[158]

This brings us to Nourse's third conception of 'judgment as critique', a model which represents an evolution of 'judgment as community' in the sense that it moves

---

[152] *Ibid*.
[153] *Ibid* at 1385.
[154] *Ibid* at 1387.
[155] Lee, *op cit*. at p.244.
[156] *Ibid* at p.235.
[157] *Ibid* at p.244. See generally, Lee, *op cit.* at Chapter 10: "The Act-Emotion Distinction", p.260ff.
[158] This appears to be the conclusion of the Scottish courts in relation to provocation: *Drury v HM Advocate* 2001 SLT 1013. However, the provocation in question was the revelation of sexual infidelity, and it is hard to imagine when it would ever be normatively acceptable to react to infidelity by killing.

beyond common sense[159] and any kind of static, community notion of normative values in order to, in all cases, 'judge anew'.[160] Judgement is thus forward-looking. Nourse argues that the corresponding dangers of mechanistic and pure character theories of emotion is that they either rely on a notion of men as beasts driven by impulse, or ask us to aspire to "standards of angelic perfection".[161] Neither option is attractive because they involve unrealistic standards. 'Judgment as critique', in contrast, involves finding a middle ground – somewhere between the good and the neutral – whereby persons are judged to a human standard. The judge must not only be critical in evaluation but also compassionate, able to imagine and identify with others with whom we may disagree.[162]

Crucially, if judgement is to mean anything more than blind adherence to the status quo, then judgements should be "made with the fear and knowledge that they may be wrong".[163] Lee offers us a more practical vision outlining what this kind of reasonableness in judgement might look like: she calls for jurors to be instructed that although it is normal to be influenced by dominant social norms, they should try not to let such norms bias their decision making.[164] If jurors are uncertain whether they are influenced by such norms, they should engage in gender-, race-, and/or sexual orientation-switching to assess their biases (a process whereby the juror imagines the exact same factual circumstances, but switches some physical attribute of the accused and victim which gives rise to a stereotype, e.g. switching the genders of the abuser and victim in domestic abuse cases).[165]

It is thus within this middle ground that we can also find a solution to the problem of representative judgment and reasonableness standards. Nourse states that our evaluative beginning must start from the middle, from those embedded in particular situations and cultures.[166] To this end, Nourse places particular emphasis on the work of other evaluative academics, like Kahan and Nussbaum, for developing this model standard which focuses on the fact that current constructs for which there is widespread approval nevertheless mask problematic normative claims. She sees such work as providing better coherence with existing law because it relies on analogies from within the law to provide legitimation.[167] If the law is willing to relax the imminence requirement to protect the

---

[159] Nourse suggests that common sense may be necessary without being sufficient: "The New Normativity", *op cit*. at 1463, fn.170.
[160] *Ibid* at 1464, citing the work of the political philosopher Hannah Arendt.
[161] *Ibid*.
[162] *Ibid*.
[163] *Ibid*.
[164] Lee, *op cit*. at pp.252-3. For a more general discussion on "judgmental descriptivism", see pp.246-54.
[165] *Ibid*. For an example of race-switching, see at p.253ff.
[166] Nourse, "The New Normativity", *op cit*. at 1465.
[167] Pace, Fingarette: see section 7.3.3 above.

honour and dignity of a man who holds his ground, why can the law not do the same for the battered woman when her honour and dignity are at stake?[168] In that sense, 'judgment as critique' essentially boils down to judgement critiquing and informing further judgment.

This necessarily involves placing responsibility on judges and juries to navigate these norms to find the right answer to each case, a solution therefore not free from error, but the criminal legal system would be honest about its intentions.[169] Better yet, it would be transparent. Transparency of reasoning would allow society to object and place pressure on the system to change when a judge does make problematic evaluations. Berger points out that this transparency might lead to the exposure of regressive and hurtful conceptions of emotional judgements that society nevertheless endorses, a point which he downplays by suggesting that the criminal law, whatever it is expressing, starts from a baseline assumption that it serves diversity, antisubordination, and equality and thus would not give countenance to such views.[170] For the most part I believe this to be true. However, even if Berger is wrong about this purpose for criminal law, it would be of little material difference. If regressive attitudes are revealed in legal judgments with which the polity agree, arguably the job is on educators to shift understandings and expectations. Further, by adopting the 'judgment as critique' approach to such questions this issue falls away as, unlike other evaluative theories, 'judgment as critique' does not appeal to some privileged sense of what is 'common' or 'typical', relying instead upon the law itself and the underlying values as a measure of what we share.[171]

7.5 Conclusion

This chapter has outlined various psychological theories of emotions in the context of criminal law defences in support of the argument advanced in chapter five that the reactive defences of coercion and necessity in Scots law should be differentiated on normative grounds: i.e. whether the actor experienced a strong emotional reaction to

---

[168] Nourse, 'The New Normativity', *op cit*. at 1465-66.

[169] The presence of a jury and thus the potential for jury independence/nullification would also add a separate layer of protection where a judge acted out of line with the social mores of the day – on the potential for jury independence as a means of redressing injustices, see generally MR Kadish & SH Kadish, *Discretion to Disobey: A Study of Lawful Departures from Legal Rules* (Stanford University Press, 1973) at pp.45-72; CS Conrad, *Jury Nullification: The Evolution of a Doctrine* (Cato Institute Press, 1998). See also SJ Schulhofer, "Harm and Punishment: A Critique of Emphasis on the Results of Conduct in the Criminal Law" (1974) 122 *University of Pennsylvania Law Review* 1497 at 1522ff. who highlights that the potential for jury nullification forces legislatures to carefully consider the grading of offences in the context of their (desired) deterrent effect.

[170] Berger, *op cit*. at 126.

[171] Nourse, "The New Normativity", *op cit*. at 1466.

stressful stimuli, referred to in this thesis as a situation of extreme pressure, or whether they sought to bring about a positive result (situation of individual emergency). Specifically, it has been necessary to outline in what way emotions can be said to provide a normative basis for exculpation in situations of extreme pressure. I have argued that, in line with the evaluative theory of emotions presented by Kahan and Nussbaum (and indeed cognitive theories more generally), emotions can be regarded as not only responsive to reason, but positively formulated by cognitive beliefs and value appraisals. This makes them suitable candidates for moral judgement in a way which undermines the current trend in criminal justice systems of treating the impact of emotions on persons as being behavioural, and hence capacity-eliminating, in nature. Under this theory, a defence based in emotional response should not be a troubling concept. It is broad enough to encompass cases where persons do not exhibit the archetypal emotional reactions we expect, and yet discriminating enough to exclude those who experience genuine but unreasonable or irrational reactions to (trivial) stimuli.

In light of this foundational basis for a defence based on emotions, I then explored the concept of 'normative judgement' to determine how one might assess the relative blameworthiness of a person who claims to act in a situation of extreme pressure, i.e. under the influence of strong emotions. It has been argued that a moderate and robust approach should be taken to the reasonableness question – one which seeks to ascertain both a subjective typicality, as well as a more objective normative reasonableness. This robust approach allows the legal system to be honest about the nature and influence of emotions in criminal law defences: they provide the basis to exculpation under situations of extreme pressure but cannot, in and of themselves, excuse behaviour which otherwise the polity would find repugnant or in breach of harm or equality principles.

Thus, and to return to the initial claim made in chapter five, coercion in Scots law should be understood as a defence based on an understandable emotional reaction to stressful stimuli (however that reaction might manifest), where the reasonableness of the conduct is determined by both the typicality of such a response, as well as the reasonableness of the response in a normative framework which upholds principles such as equality and freedom from harm. As a result, any requirements should adhere to this premise. Necessity should correspond to situations of individual emergency, where the presence of a constrained choice should also be evaluated on normative principles, but here based on the idea that a person acted virtuously to bring about a net positive result (whether or not they were successful in this goal). The assessment is therefore one of the

circumstances surrounding the *prima facie* criminal act. In part three of this thesis, I re-examine two of the more problematic requirements of both coercion and necessity in contemporary Scots law in light of this new framework: the requirements that a threat must be immediate, and that it relate to serious injury or death.

PART III

8. Temporal Requirements in Reactive Defences

8.1 Introduction

Part II of this thesis examined the rationale for reactive defences like coercion and necessity in criminal law. It was argued that these defences stem from a normative foundation which recognises a legal distinction between persons who freely commit offences, and those who do so only because their choices or will were constrained. These constraints are not to be regarded as involuntary: indeed, even where the will is constrained and a person acts under extreme emotional pressure, they can still be said to be responsible for their conduct, based on an underlying value system to which they subscribe, and on which it is reasonable for the law to pass normative judgement. Part III of this thesis, beginning with this chapter, will look at some of the more controversial requirements of the coercion and necessity defences in Scots law (originally outlined in Part I of this thesis) in the context of this new conceptual understanding to determine their rationality and, ultimately, the desirability of any future reforms.

This chapter will examine temporal requirements in reactive defences generally, through an extensive examination of the debate surrounding the rationale for an imminence requirement in self-defence, to determine the rationality of temporal requirements in both situations of individual emergency and extreme pressure. Understood as a justification defence, the imminence of harm requirement is of particular importance in self-defence owing to the nature of the harm (potentially lethal violence) being endorsed. This has generated a rich discussion about the precise nature and limits of when it is permissible to defend oneself from a violent aggressor. With the normative foundations of reactive defences explored here being potentially broad enough to encompass self-defence,[1] this chapter explores the moral and political theories relied on to rationalise temporal requirements in the self-defence (i.e. justificatory) context, being mindful of important distinctions in other reactive defences such as the identity of the victim and the nature of the harm committed, to draw an important distinction between a general necessity

[1] See section 5.1.

requirement[2] on the one hand, and a more specific inevitability of harm requirement for justified conduct on the other. This distinction will allow us to consider both principles of necessity and inevitability of harm in other reactive defences to reach the conclusion that a strict temporal requirement, while having probative value, should not be a necessary requirement for any reactive defence which is not justificatory in nature, such as situations of extreme pressure and some situations of individual emergency.

A note should be made on the terminology used throughout this chapter, in relation to the rest of this thesis. In Scots law, the temporal requirement in the necessity and coercion defences is understood by reference to the terms 'immediate' or 'immediacy'. Thus, there must be an 'immediate danger' of harm or an accused cannot plead this type of reactive defence. In contrast, the temporal requirement in self-defence is described by reference to the term 'imminence'. At a descriptive level, this could suggest that the temporal requirement in self-defence is less strict than in necessity and coercion, with immediacy implying contemporaneousness between the threatened harm and the offence committed to avoid it. In practice, however, the treatment of temporal requirements for both self-defence and necessity/coercion in Scots law appears to be essentially the same.[3] Nevertheless, and for clarification purposes, in the ensuing argument the term 'imminence' and its derivations are used exclusively to refer to the temporal requirement in self-defence, with the term 'immediate' and its derivations being used exclusively to refer to the temporal test in necessity/coercion.

8.2 Temporal Requirements in Reactive Defences: An Overview

As discussed in chapter four, one of the core features of reactive defences such as necessity and coercion in Scots law is the requirement that the danger which the accused faced share a close temporal connection with the criminal act that the accused committed to avoid it. If the threatened harm to the accused exists at some point in the future, so the argument goes, it becomes more likely that the accused had alternative lawful courses of action available to them, such as seeking protection from public authorities. Some question the underlying efficacy of this rule, arguing that although elements may be more or less

[2] In this chapter, references will be made to the 'necessity principle' as a guiding rationale which informs the requirements of self-defence. This is to be distinguished from the substantive defence of necessity referred to elsewhere in this thesis. Unless stated otherwise, references to necessity in this chapter refer to the principle, and not the substantive defence.

[3] At least no case exists discussing their relationship or any distinction between them. For an overview of temporal requirements in necessity/coercion and self-defence, see sections 4.2 and 8.3.1 respectively.

'likely', the fact remains that in some circumstances there may be a valid threat which dominates the mind of the accused and for which no reasonable alternative course of action exists, despite a loose temporal connection between the threat and the offence.[4]

This thesis has argued that the coercion and necessity defences in Scots law should be understood via a different distinction – one which recognises a difference between a constrained choice and a constrained will – under the headings of situations of individual emergency, and situations of extreme pressure. Thus, requirements like contemporaneity between the threat and offence require two separate analyses for each rationale. Where the rationale for exculpation rests on the fact that a person's will was compromised by the presence of strong emotions, such that it informed their reasons for acting as they did (situations of extreme pressure), the focus is on establishing factors which would cause such a deterioration of ordinary will. A close temporal connection may strengthen a claim that such a deterioration occurred, but it is unlikely to be necessary in all cases to demonstrate this. In contrast, and where the rationale for exculpating is based on a lack of available alternative choices where an actor attempts to minimise harm (situations of individual emergency), there is a greater focus on the pressure and limits of the situation itself, such that it would have caused an ordinary person to have acted as the actor did. There is therefore a greater emphasis on proving a lack of reasonable alternatives such that an emergency indeed arose, and a close temporal connection will be an effective indicator of this.

Temporal connections have not received much attention in relation to the coercion/duress and necessity defences in legal literature:[5] this is mostly due to a lack of reported cases.[6] In contrast, literature on the imminence requirement in self-defence is voluminous, and thus provides a rich analytical tool with which to frame the issue.[7] I shall

---

[4] Discussed at section 8.3.1 below.

[5] Discussions in the literature appear to be limited to the context of international (criminal) law, specifically state justifications for war and the resulting criminal liability for individuals accused of war crimes. See e.g.: J McMahan, *Killing in War* (Oxford University Press, 2009), at pp. 59 and 71; J Bond & M Fougere, "Omnipresent Threats: A Comment on the Defence of Duress in International Criminal Law" (2014) 14 *International Criminal Law Review* 471. An attempt has even been made to synthesise the self-defence rules in international law with those of domestic criminal law: S Wallace, "Beyond Imminence: Evolving International Law and Battered Women's Right to Self-Defense" (2004) 71 *The University of Chicago Law Review* 1749.

[6] *R v Ruzic* [2001] 1 SCR 687 represents one of the few which is directly concerned with this point.

[7] Provocation is another reactive defence in Scots law which features a temporal requirement, but its status as a partial defence makes any analogy less salient. See *Cochrane v HM Advocate* 2001 SCCR 655 at 668ff. (670 in particular) considering and rejecting the view of Lord Lane in *R v Graham* [1982] 1 WLR 294 at 300 that "a common approach falls to be applied in cases of duress and provocation". Lord Justice-General Rodger was sceptical of a connection between these defences for several reasons, including the fact that provocation is only a partial defence (para [25], 669). This is despite noting similarities between both necessity (para [17], 666) and self-defence (para [20], 667). See also Lord Hailsham's comments on any

therefore outline the controversy surrounding this requirement in self-defence as a way of then exploring the issue in relation to reactive defences more generally. Indeed, both situations of extreme pressure and individual emergency feature similar aspects to those in self-defence. In particular, all three scenarios involve the actor being placed in a stressful situation where they are presented with a threat of great harm, with no reasonable alternatives other than suffering the harmful consequences or committing a *prima facie* crime to escape them.

The primary distinction from other reactive defences is that in self-defence cases, the source of the harm and the victim are the same person (the aggressor). This is not an insignificant difference, but it is of less importance in the context of temporal requirements, which are primarily understood as being relevant to determining the authenticity of the situation giving rise to mitigating circumstances, rather than the reasonableness of the consequences of the act. Indeed, this distinction may provide a unique way of framing the temporal issue: self-defence inherently features competing principles which encapsulate not just the act of the accused, but also the initial aggressive act of the victim and the context in which the need for defensive force arises. Thus, and for example, despite the necessity principle seemingly demanding a duty to retreat where safe to do so,[8] the rules on self-defence in some jurisdictions do not require this, suggesting that principles other than necessity are at play.[9] Discovering these principles and determining their relevance to other reactive defences may therefore provide fresh insight.

In addition, self-defence is understood as a justificatory defence. When a person kills another in self-defence they are not just blameless: the law states that the action was permissible, or even right. In contrast, this thesis has argued that necessity and coercion in Scots law, insofar as they are recognised and applied in cases, are more likely to be understood as excusatory in nature, meaning that the actor is blameless but the action wrong.[10] In theory, however, it is possible to conceive of these defences in justificatory terms and, in fact, some jurisdictions do understand necessity as a form of lesser evils

---

comparison between duress and provocation in *R v Howe* [1987] AC 417 at 435, Lord Griffiths at 445, and Lord Mackay at 455-6.

[8] See e.g. M Dsouza, "Retreat, Submission, and the Private Use of Force" (2015) 35 *Oxford Journal of Legal Studies* 727. Dsouza argues that a duty to retreat necessarily flows from the availability of alternate courses of action than the private use of force.

[9] Cf. D Akande & T Liefländer, "Clarifying Necessity, Imminence, and Proportionality in the Law of Self-Defense" (2013) 107(3) *American Journal of International Law* 563 at 564 questioning the precise requirements of necessity: "[I]s an active exhaustion of peaceful alternatives necessary, or does the mere appreciation that an alternative is not equally effective suffice?".

[10] See above at section 4.5.

utilitarianism which operates on the basis that the conduct is justified.[11] Likewise, and despite self-defence being regarded as the archetypical justification defence in Scots law, it is not difficult to conceive of self-defence as an excuse.[12] For example, Fletcher argues that reasonable mistakes about the grounds for self-defence should only ever excuse an accused, not justify their conduct.[13]

Finally, this thesis has developed the concept of situations of individual emergency on the basis that the actor intends to minimise harm, irrespective of whether they then achieve this goal. The reader will recall from chapter five that the normative benchmark was set at understandable conduct,[14] although what counts as understandable will depend on the underlying rationale (i.e. constrained choice vs constrained will). Thus, while some aspects of the self-defence analogy may be of limited relevance to situations of extreme pressure due to the differing nature of exculpation,[15] it is argued that a stronger analogy is to be found in relation to situations of individual emergency, particularly those where the actor's conduct does end up being justified.

8.3 Imminence in Self-defence

As aforementioned, an analysis of the imminence requirement in the law of self-defence can be utilised to develop a better understanding of the rationale for temporal requirements in reactive defences generally. Like other reactive defences, circumstances can arise which seem to warrant a valid plea of self-defence but for the lack of an 'immediate threat'. Specifically, a great deal has been written about the imminence requirement in self-defence owing to the controversy surrounding its potential to negate the plea in cases of domestic abuse involving nonconfrontational killings, such as killing a sleeping abuser. The domestic abuse scenario also engages rationales which are not justificatory in nature: 'battered woman syndrome' is seen as a pathology which seeks to

---

[11] See, e.g. the US Model Penal Code, §3.02 Justification Generally: Choice of Evils. Indeed, Dubber refers to necessity as "the fallback justification… if others fail": MD Dubber, *An Introduction to the Model Penal Code* (Oxford University Press, 2015) at p.146.

[12] For arguments on self-defence understood as an excuse, see: CO Finkelstein, "Self-Defense as a Rational Excuse" (1996) 57 *University of Pittsburgh Law Review* 621; CJ Rosen, "The Excuse of Self-Defense: Correcting A Historical Accident on Behalf of Battered Women Who Kill" (1986) 36 *American University Law Review* 11.

[13] See GP Fletcher, "Domination in the Theory of Justification and Excuse" (1996) 57 *University of Pittsburgh Law Review* 553 generally.

[14] See above at section 5.2.1.

[15] Although cf. Dressler, who argues that women who defend themselves from their abusive partners should be granted a no-fair-opportunity duress claim: J Dressler, "Battered Women and Sleeping Abusers: Some Reflections" (2006) 3 *Ohio State Journal of Criminal Law* 457 at 470

explain the accused's conduct, rather than endorse their act. These discussions may therefore be illuminating for reactive defences generally. Broadly speaking, this section introduces the concept of inevitability and examines its interaction with the competing principle of necessity as understood through the law of self-defence. Once established, this concept of inevitability is then examined in greater detail in the following section to determine precisely how it justifies an accused's use of force.

8.3.1 An overview of the imminence requirement in self-defence

The concept of self-defence[16] permits a person to use (lethal) force against an aggressor to protect themselves or others from unlawful harm. Scots law only permits the use of self-defensive force where the accused reasonably believes that they were in imminent danger.[17] In a murder case, the accused must fear that someone (either the accused or a third party) will be killed or suffer serious injury imminently unless they act.[18] This requirement has been highly controversial over the last thirty years, owing in large part to the emergence of discussions which consider whether a person who kills their abusive partner can claim self-defence.[19] Any person that kills another in the course of fending off a life-threatening physical attack from that person will *prima facie* meet the requirements necessary to establish the defence. However, some[20] victims of abuse end up

[16] This concept includes the defence of others. The term 'private defence' is sometimes used as an umbrella term for both.
[17] *Owens v HM Advocate* 1946 JC 119.
[18] *HM Advocate v Greig*, Unreported, High Court, May 1979, cited in: P Ferguson & C McDiarmid, *Scots Criminal Law: A Critical Analysis* (2nd edn, 2014), p. 561.
[19] The scholarship is too immense to list fully, but see e.g.: MJ Willoughby, "Rendering Each Woman Her Due: Can a Battered Woman Claim Self-Defense When She Kills Her Sleeping Batterer" (1989) 38(1) *University of Kansas Law Review* 169; A McColgan, "In Defence of Battered Women Who Kill (1993) 13(4) *Oxford Journal of Legal Studies* 508; R Rosen, "On Self-Defense, Imminence, and Women Who Kill Their Batterers" (1993) 71(2) *North Carolina Law Review* 371; Fletcher, "Domination", *op cit*. (1996); JB Murdoch, "Is Imminence Really Necessity? Reconciling Traditional Self-Defense Doctrine With The Battered Woman Syndrome" (2000) 20(1) *Northern Illinois University Law Review* 191; AS Burke, "Rational Actors, Self-Defense, and Duress: Making Sense, Not Syndromes, out of the Battered Woman" (2002) 81(1) *North Carolina Law Review* 211; KK Ferzan, "Defending Imminence: From Battered Women to Iraq" (2004) 46(2) *Arizona Law Review* 213; S Wallace, *op cit.* (2004); JC Moriarty, "While Dangers Gather: The Bush Preemption Doctrine, Battered Women, Imminence, and Anticipatory Self-Defense" (2005) 30(1) *New York University Review of Law & Social Change* 1; F Leverick, *Killing in Self Defence* (Oxford University Press, 2006) at 102ff.; WRP Kaufman, "Self-Defense, Imminence, and the Battered Woman" (2007) 10(3) *New Criminal Law Review: An International and Interdisciplinary Journal* 342; A Guz & M McMahon, "Is Imminence Still Necessary? Current Approaches to Imminence in the Laws Governing Self-Defence in Australia" (2011) 13(2) *Flinders Law Journal* 79; M Baron, "Self-Defense: The Imminence Requirement" in L Green & B Leiter (eds), *Oxford Studies in Philosophy of Law: Volume 1* (Oxford University Press, 2011) pp.228-266; S Goosen, "Battered Women and the Requirement of Imminence in Self-Defence" (2013) 16(1) *Potchefstroom Electronic Law Journal* 70.
[20] There is a disagreement here about the precise numbers of those who kill their partners in response to violent confrontations versus nonconfrontational situations where the abusive partner is calm or even sleeping. Historically, conventional wisdom held that most women killed in nonconfrontational

killing their abusive partners in nonconfrontational situations, such as when their partners are sitting watching the television[21] or even asleep.[22] Conditions for the accused leading up to such killings are often described as analogous to those held hostage,[23] making traditional escape routes unrealistic.[24] Under traditional self-defence rules, those persons are denied a defence on the basis that they were not acting to avert any imminent threat of harm. This conclusion has led to a backlash from legal commentators, with some calling for either a complete or partial abolition of the temporal requirement in self-defence.[25]

Abolitionist approaches tend to focus on the idea that purported self-defensive action should be considered in the 'totality of the circumstances'.[26] The foundation for such claims stems from the belief that the imminence requirement is indivisible from a broader necessity principle, which serves to establish that force was necessary. Thus, when we seek to find a close temporal connection between an accused's act and the act of aggression, we are in reality seeking to determine whether the self-defensive force was

---

circumstances: R Rosen provides an extensive list of examples, *op cit*. at 402 fn. 80. However, more recently this view has been challenged, with Maguigan suggesting that only a very small minority (20%) are killed in such circumstances: H Maguigan, "Battered Women and Self-Defense: Myths and Misconceptions in Current Reform Proposals (1991) 140(2) *University of Pennsylvania Law Review* 379 at 397. It should be noted that Maguigan's numbers came from counting the number of appellate court decisions published between 1902 and 1991; it therefore does not account for those homicides that did not result in dismissal prior to trial or those resolved by guilty pleas – a point made by Dressler (although also recognised by Maguigan) *op cit*. at fn.1. See also LL Ammons, "Discretionary Justice: A Legal and Policy Analysis of a Governor's Use of the Clemency Power in the Case of Incarcerated Battered Women" (1994) 3(1) *Journal of Law and Policy* 1 at 57, fn.209 stating that when reviewing 123 files of female inmates in Ohio petitioning for clemency based on claims of self-defence in domestic violence deaths, 92 of the homicides occurred while the victim was "awake, alert and beating the woman". Wallace claims that the imminence requirement is problematic even in confrontational cases: *op cit*. at 1760-1. More recently, McPherson analysed 62 cases in which a female was accused of killing her abusive partner in Scotland between 1990 and 2018. She found that seven of the homicides occurred in nonconfrontational circumstances: R McPherson, "Legal change and legal inertia: understanding and contextualising Scottish cases in which women kill their abusers" (2021) 5 *Journal of Gender-Based Violence* 289 at 297.

[21] See, e.g. the Australian case *Collingburn* (1985) A Crim R 294.

[22] Such as in the infamous US case of *State v Norman* 324 NC 253, 378 S.E.2d 8 (1989) in North Carolina where Judy Norman shot her husband in the head while he slept. The deceased abused Judy for years, consisting of both physical and psychological harm, including forcing her to eat dog food and to sleep at the foot of their bed.

[23] Willoughby, *op cit*. at 180 and again at 185 (comparing battered women to hostages); M Mahoney, "Legal Images of Battered Women: Redefining the Issue of Separation (1991) 90(1) *Michigan Law Review* 1 at 87 (discussing the US case of *State v Hundley*, 693 P.2d 475 (Kan. 1985) in which the hostage analogy was raised) and again at 92; M Dowd, "Dispelling the Myths About the 'Battered Woman's Defense:' Towards a New Understanding" (1992) 19 *Fordham Urban Law Journal* 567 at 580; McColgan, *op cit.* at 519; SD Appel, "Beyond Self-Defense: The Use of Battered Woman Syndrome in Duress Defences" (1994) 4 *University of Illinois Law Review* 955 at 976; Guz & McMahon, *op cit.* at 119; R *v Lavallee* [1990] 1 SCR 852 at 889 per Wilson J (Canada); *State v Hundley*, 693 P.2d 475, 479 (Kan. 1985) at 467 (US); *State v Johnson* 956 F.2d 894, 899 (9th Cir. 1992) at 899 (US).

[24] Mahoney, *ibid* at 87: "the persuasive power of the hostage analogy depends on the recognition that the woman in an abusive relationship is not free to leave".

[25] Others advocate for a new, bespoke defence to deal with abusive partner homicides, although cf. Baron, *op cit*. at 235.

[26] Moriarty, *op cit*. at 25.

necessary.[27] Situations where a woman kills her abusive partner in nonconfrontational circumstances may therefore suggest that a strict temporal requirement is unprincipled under this theory, because it would deny a defence even when the necessity principle points to there being one.[28] In other words, the limits of an imminence requirement should be dictated by the necessity principle.[29] Indeed, where the accused kills in the context of a one-off adversarial encounter, a strict temporal requirement is appropriate for demonstrating a lack of alternative options – but in other cases such as hostage scenarios, the use of force may be manifestly reasonable despite the absence of an imminent physical threat.[30] The fact that the accused had time to consider their options before responding to the threat should, in every case, affect the credibility of their claim that defensive force was necessary, but it should not disqualify the plea outright without further inspection.[31]

With the above in mind, Burke suggests that the accused should be permitted to explain to the jury the circumstances surrounding the abusive relationship in order to build a defence case from which they claim the use of force was necessary.[32] Likewise, Moriarty advocates for examining the surrounding circumstances in self-defence claims, arguing that the current focus on singular moments has created an "overly rigid application of the imminence requirement" and thus that courts should examine a broader spectrum of time and context in which the killing occurred when evaluating a self-defence claim.[33] Both of these accounts broadly follow the same logic – if juries and courts knew more about the plight of the domestic abuse victim, there would be no need for an imminence requirement to act as sole proxy for the existence of necessity, with juries able to appropriately assess the legitimacy of the self-defence claim on other grounds.

McColgan emphasises this fact, pointing out that a jury's assessment of a woman's belief that she was under threat of attack, and of the seriousness of an anticipated attack, "will clearly be influenced by evidence of the abuser's past conduct".[34] Thus, it is important to considerations of fairness and justice that such evidence is not excluded or

---

[27] R Rosen, *op cit*. at 380; Baron, *op cit*. at 233-4.
[28] Willoughby, *op* cit. at 181. Although not every battered woman: see Burke, *op cit.* at 274 and 295-7. Cf. Dressler, *supra* fn.14.
[29] F Allhoff suggests that cases in which a plea of self-defence was defeated due to a lack of imminence could equally have come to the same conclusion based on a lack of necessity – given that the latter is more robust for dealing with difficult cases (such as battered women) he argues in favour of abrogating imminence for necessity: "Self-defense Without Imminence" (2019) 56 *American Criminal Law Review* 1527, in particular at 1541ff.
[30] McColgan, *op cit*. at 518.
[31] Allhoff, *op cit*. at 1542ff.
[32] Burke, *op cit*. at 316.
[33] Moriarty, *op cit*. at 2.
[34] McColgan, *op cit*. at 528.

minimised from the factual narrative and, in light of these considerations, the imminence requirement should be removed insofar as it acts as an obstacle to the development of that narrative. However, in most other cases the imminence requirement is an incredibly effective way of demonstrating the required necessity, and so Rosen suggests a partial abolition, or rather exception, for self-defence cases where evidence of domestic abuse is raised by the accused: in such circumstances the jury would instead determine whether the accused's use of deadly force was necessary, based on their reasonable beliefs in the circumstances.[35] Other accounts suggest reinterpreting the imminence requirement in the context of domestic abuse cases, such that the requirement should be considered to be automatically met on the basis that 'battered women',[36] due to their heightened and perpetual state of anxiety, are constantly under an imminent fear of harm or death.[37]

The necessity principle, however, is not the only potential basis for the imminence requirement in self-defence, and thus there has been some scepticism as to whether this account provides an accurate representation of the requirement's rationale. For those who argue against the 'imminence as necessity' thesis, such accounts fail to consider a crucial point about self-defence which sets it aside from other reactive defences which, they argue, rely more heavily on the necessity principle than does self-defence. In this context, authors can therefore be said to fall into one of two camps: those, outlined above, that believe the imminence requirement is a proxy[38] or 'translator'[39] for a broader necessity principle;[40] and those who argue that the requirement serves a different master – inevitability.

To mark this conflict, I identify two theories. The 'imminence as necessity' theory makes the following broad claim – if one believes that there is a connection between imminence of harm and necessity such that the former cannot be understood independently from the latter,[41] it necessarily follows that the imminence requirement should adhere to

---

[35] R Rosen, *op cit.* at 404-5.
[36] It is unclear whether this analysis is transferable to scenarios where the abused partner is male, and indeed whether such an approach would demand a different set of rules depending on the respective gender identities of the abused and abusive partners, but an in-depth discussion of this point goes beyond the scope of what is being argued here.
[37] Leverick sets out several of these views: *op cit*. at 94-6. See also Willoughby, *op cit*. at 182.
[38] Allhoff, *op cit*. at 1542.
[39] R Rosen, *op cit*.
[40] See also Murdoch, *op cit*. In the context of duress in English and US law, see J Dressler, "Duress" in J Deigh & D Dolinko, *The Oxford Handbook of Philosophy of Criminal Law* (Oxford University Press, 2011), pp.269-298 at 271.
[41] R Rosen, *op cit.* at 380; Allhoff, arguing that imminence and necessity "travel together": *op cit.* at 1531; Burke, *op cit.* at 241; Murdoch, *op cit*. at 212: imminence is "merely a way of measuring necessity"; Guz & McMahon, *op cit.* at 97. So too, in the context of international law is imminence in self-defence often treated as an indicator or proxy for necessity: Akande & Liefländer, *op cit.* at 565: "The better argument is that where a threat is sufficiently probable and severe, the mere fact that it is still temporally remote should provide no independent injunction against action where that action is necessary and proportionate"; D Rodin,

the necessity principle and must either be relaxed or abolished to accommodate for those instances where necessity, but not imminence of harm, exists.[42] In that sense the claim is one that the temporal requirement in self-defence is merely an evidential matter, serving to lend credibility to a claim that defensive force was necessary.[43] In contrast, academics who adhere to the alternative 'imminence as inevitability' theory tend to gravitate towards a conclusion whereby the imminence requirement is retained based on varying explanations for its independent role from necessity, all of which I shall argue feature the same underlying concept of inevitability.

8.3.2 Inevitable harm

The 'imminence as inevitability' theory requires additional exposition because, unlike the necessity principle, no such broad, universally accepted principle of inevitability exists in the current literature.[44] I shall therefore develop this concept in the context of the 'imminence as inevitability' theory through an analysis of some of the more prominent accounts which, in my view, can be said to adhere to the principle that harm must be inevitable. To be clear, the concept of inevitability as developed here should be construed broadly in line with the word's natural meaning: if the subject (i.e. threatened harm) can be described as so probable that it is certain to happen or unavoidable (but for the accused's intervention), then we can say that it is inevitable. Thus, imminence as inevitability theories essentially claim that temporal requirements operate to prove, sufficient for a court of law, that the harm was certain to happen and thus the accused's intervention was warranted.

Leverick provides the most explicit account of inevitability when she states that what is important to self-defence in the context of a temporal requirement is determining

---

War and Self-Defense (Oxford: Clarendon Press, 2002) at 41: "Imminence, like the duty to retreat, is simply a component and corollary of the requirement of necessity".

[42] The Model Penal Code adopts this theory by eliminating the imminence requirement and replacing it with a singular requirement that defensive force be "immediately necessary": §3.04(1).

[43] Or reasonable, as in Canada: *R v Lavallee* [1990] 1 SCR 852 at [24], approved in *R v Pétel* [1994] 1 SCR 3. These cases helped shape the new statutory provision on self-defence, s34 of the Canadian Criminal Code, which includes imminence as one of many factors in determining whether the act was reasonable in the circumstances: s34(2)(b).

[44] Although it is not a novel concept: see, F Leverick, *op cit.* at p.101ff outlining an 'inevitability of harm' test. Baron comes close to recognising the importance of inevitability when she correctly identifies the common mistake of conflating the literal definition of 'imminence', with an understanding of imminence that it "convey[s] a high degree of certainty that the dreaded thing will happen unless it is stopped": Baron, *op cit*. at 248-9.

the point at which the probability of harm is sufficient to allow such claims.[45] In her opinion the level should be set at, or very close to, inevitability, "exactly the sort of level signified by an imminence requirement".[46] Imminence of harm is important says Leverick because, as a central premise, the law should be concerned with respecting the right to life of all people, and any relaxation of the imminence requirement would lower the probability that the threat would actually materialise, thereby increasing the likelihood that killing the (potential) aggressor was not necessary at all.[47] The suggestion is that necessity, as a guiding principle, pays insufficient regard to the aggressor's right to life, and may encourage unnecessary killing because it can be satisfied before a choice *must* be made between the lives of the aggressor and defender.

Other accounts which support this idea, that the necessity principle alone pays insufficient regard to the value of the victim's life in self-defence, do so by incorporating ethical or political considerations to highlight the importance of an inevitability of harm requirement to the justification of using lethal force. Ferzan, for example, has argued that the imminence requirement corresponds not to the necessity principle, but to an initial act of aggression by the victim which gives rise to the claim of self-defence.[48] This is so, says Ferzan, because what is important to establish in self-defence, and indeed what distinguishes it from other defences of self-preference, is the type of threats that trigger the right to retaliate.[49] Acts of aggression trigger such rights, according to Ferzan, because the stipulation of a threatening act requires not just indications of intent but also a *right* of response, and therefore self-defence is only understandable as a response to another's aggressive conduct.[50]

Thus, in a scenario somewhat analogous to offences of strict liability, the imminence requirement provides the '*actus reus*' for aggression, so claims Ferzan, in a way congruent with the criminal law's approach to incomplete attempts.[51] Under Ferzan's theory of

---

[45] Leverick, *ibid* at p.101-2. She recognises that the New Zealand Law Commission also reached a similar conclusion in its Report (*ibid* at 102 fn.89): *Some Criminal Defences with Particular Reference to Battered Defendants* (R73, 22 May) (Wellington: NZLC, 2001) at 12, paras 30-32.
[46] *Ibid* at p.102.
[47] *Ibid* at p.101. See also O Bakircioglu, *Self-Defence in International and Criminal Law: The Doctrine of Imminence* (Routledge, 2011) at p.237.
[48] Ferzan, *op cit*., in particular at 255ff.
[49] *Ibid*.
[50] *Ibid* at 257.
[51] *Ibid* at 257-8. The term 'incomplete attempts' appears to map to the concept of a 'non-last act' attempt: G Yaffe, *Attempts* (Oxford University Press, 2010) at 25-27 (referring to the distinction between 'last act' and 'non-last act' attempts); L Alexander & KK Ferzan, *Reflections on Crime and Culpability: Problems and Puzzles* (Cambridge University Press, 2018) at 6: "An incomplete attempt is a step leading to what the actor intends will culminate in an act that he believes will be culpable." For Scots law see Gordon, *Criminal Law*, at para. 6.04ff. An argument can be made that, in the context of self-defence where what is required,

imminence in self-defence, the victim's act of aggression provides the accused with a moral justification for utilising defensive force. Where an accused applies force *before* there is an imminent threat, that force cannot be said to have obtained any defensive quality and the accused therefore acts with no moral high ground (i.e., without justification). An act of aggression by the victim, i.e. a confirmed rather than speculative attack, ensures that this moral threshold is met. What is important is that the victim reaches a point of no return where their actions trigger a moral judgement giving rise to the right to respond. In this sense, Ferzan's theory views imminence of harm as a demonstration of the inevitable, or as close as can be reasonably achieved in a world where the attack is eventually unsuccessful.

There are certain elements of Ferzan's account which are less convincing. Specifically, her understanding of acts of aggression as being characterised by an underlying 'guilty act'[52] appears to create two different classes of aggressor: culpable aggressors, to be contrasted with non-culpable aggressors. Indeed, if what justifies self-defence is the existence of a prior moral transgression by the victim, Ferzan's theory struggles to explain why self-defence is permitted in cases where the aggressor lacks capacity, and thus culpability (e.g., children or those with mental disorders). This is because such attacks would not carry the level of guilt required to empower the accused to retaliate with defensive force, ergo rejecting Ferzan's theory. However, the broader implications for the imminence requirement in self-defence are important. The temporal requirement is seen as an indicator of an inevitable harm, and it is only where there is inevitable harm (characterised here as an unlawful attack which generates a moral right to respond) that lethal force may be deployed.

Kaufman has likewise equated the imminence requirement in self-defence with a requirement for an objective threat of inevitable harm. He rejects the imminence as necessity theory because it fails to recognise the independent role of the imminence requirement which he sees as distinguishing between the permissibility of citizens and the state to use force.[53] In ordinary circumstances, the state has a monopoly on the use of force and may use such force where necessary to maintain the legitimate aims of the legal system, even when there is no imminent threat. This power is vested in the state as part of

---

according to Ferzan, is an act of aggression (and presumably a culpable one), only a complete or last act attempt could constitute the necessary aggression to satisfy the imminence requirement in self-defence. However, as the distinction is of minimal importance to the present argument, no more shall be said here.

[52] To continue the offence/defence analogy.

[53] Kaufman, *op cit*. at 351.

its role in ensuring the safety (both domestic and international) of its citizens.[54] In contrast, citizens are only authorised to use force where necessary to ward off an imminent threat. This limited exception exists because, in such situations, the state will be unable to fulfil its obligation to protect them.[55] In all other circumstances, citizens are expected to seek out the protection of authorities.[56] The basis for Kaufman's theory of an imminence requirement is therefore political rather than moral; the accused's right to defend themselves stems not from some moral high ground that they obtain, or that the initial aggressor loses, when there is an inevitable unlawful attack. Rather, it stems from the limits of public authority and the practical reality that, although ordinarily only the state is authorised to use force, at the point of an imminent attack the state would not be able to intervene in time to protect the accused, and thus the accused may respond with force as they are not expected to "submit passively to self-destruction".[57]

A theory of the imminence requirement based on public authority can therefore be understood in terms of self-defence operating as the state's delegation of force, or the polity's reservation of the right to respond to unlawful force,[58] justified on the basis that state help would be ineffective. In the context of cases where the imminence requirement has been described as problematic (i.e. the homicide of domestic abusers), Kaufman states that it logically follows that if the requirement is premised on the state's ability to otherwise provide protection against violence, then in any circumstances that the state is unable to do so the requirement should be suspended.[59] However, he seriously doubts that persons would be able to reach the 'state of nature' required, based on a complete failure of

---

[54] *Ibid* at 354-59. See also Goosen, *op cit*. at 92-4.

[55] *Ibid* at 351-2. Kaufman refers to such accounts as invoking the "Public Authority" restriction on the use of force (at 354). This view is based on Max Weber's assertion that the mark of the state is its successful claim of a monopoly on the legitimate use of physical force within a territory (M Weber & J Winckelmann, *Gesammelte Politische Schriften*, (5th edn, Mohr, 1988), at pp.505-6), as well as Thomas Hobbes' comments that the obligation of citizens to their government lasts "as long, and no longer, than the power lasteth, by which [the state] is able to protect them" (T Hobbes, *Leviathan* (Andrew Crooke, 1651), Ch. XXI, para. 21).

[56] Kaufman explains that this allocation of authority on the use of force (preemptive, punitive and restorative, on the one hand, and self-defensive on the other) stems from a natural law right of an individual to defend themselves, but this right does not extend to seeking vengeance as this is the province of a state's court system. Kaufman explains that, throughout history, an understanding of the natural law basis for state power has emerged to justify their preemptive and vengeful actions against other states and individuals, being based on the state as an objective and disinterested party, while denying the same authority to citizens on the basis that they have the ability to seek redress of their rights from the tribunals of this objective and disinterested state: Kaufman, *op cit*. 355-59. See also Gauthier, *op cit.* at 616 and 618.

[57] Kaufman, *ibid* at 354. See also Wallace, *op cit*. at 1760.

[58] Whether or not this is a delegation by the state or reservation by the polity is contested. Thorburn points to a delegation of authority: see M Thorburn, "Justifications, Powers and Authority" (2008) 117 *Yale Law Journal* 1070, generally and also at 1127. Fletcher appears to have initially supported this view (GP Fletcher, *Rethinking Criminal Law* (Oxford University Press, 2000) at p.764), but changed his views in later writings: Fletcher, "Domination", *op cit.* at 570. Kaufman believes that the right to self-defensive force is never given up to the state: *ibid* at 354, fn.33.

[59] *Ibid* at 361.

the legal system in question, other than in genuine hostage situations or contexts of systemic exclusion (e.g. the experience of the Jewish population in Nazi Germany).[60] The effectiveness of state protection is therefore (and perhaps rather naively)[61] to be construed objectively.

8.3.5 Beyond the moral and political: inevitability as a counter-axis to necessity

Despite the contrasting foundations of the political approach of Kaufman and the moral approach of Ferzan, we can once again see that the concept of inevitability provides the underlying reasoning of Kaufman's account: if an attack is not inevitable then an accused cannot rely on self-defence and must seek refuge from the authorities. Put differently, it is only when an attack is inevitable that one can be certain of the ineffectiveness of state help and can therefore resort to retaliatory force.[62] All of this is to say that, ethical or political, the basis for an imminence requirement in both accounts remains the same: there must be an inevitable attack, understood here as an attempt to inflict unlawful harm, which can be evidenced by an act of aggression to which the accused must respond. This kind of approach therefore leaves little room for a necessity principle to operate in the context of the temporal requirement. If the focus of the imminence requirement is, as suggested in each theory, states of affairs created by the act of the victim, it becomes irrelevant to consider the actions of the accused.

This is, of course, not to say that considerations of necessity are unimportant to other aspects of self-defence. Jurisdictions may (and usually do) require, along with a temporal requirement, some degree of proportionality as well as a general necessity principle which operates independently to require that, in addition to imminence of harm, no reasonable alternative course of action was open to the accused. In that sense there may well be a discussion about what options the accused had, assessed at a level ranging from the objective, reasonable person, all the way to the subjective accused in the particular circumstances. Thus, necessity is still necessary to such theories for the justification of self-defence, but it is not sufficient. Rather, approaches such as Ferzan and Kaufman's suggest that there is a clear distinction between the necessity of the situation, as perceived

---

[60] *Ibid* at 361-64. He argues that statements to the effect that imminence should be suspended when state protection is not effective are open to wide interpretation and would risk overly complicating jury trials and endless debates about definitions of 'effective'.
[61] Baron, *op cit*. at 257.
[62] Fletcher states that from an objective standpoint whether an attack is imminent will depend exclusively on the qualities of that attack – namely its proximity to success and on how much harm it is forecast to cause: Fletcher, "Domination" *op cit*. at 570.

by the accused, and the probability of the harm, as presented to the accused by the victim.[63] In other words, the necessity principle alone cannot justify self-defensive action (although it might excuse it); for such acts to be acceptable the law requires an objective inevitability of harm.

8.4 The Role of Inevitability in the Justification Calculation

The previous section explored different theories in support of a temporal requirement independent of necessity and discovered that, in their rejection of the necessity principle as providing the underlying rationale for the imminence requirement in self-defence, both ethical and political accounts could be described as theories which recognise this requirement as demonstrating inevitability. The concept of inevitability, it was argued, is essential to the classification of an act as justified because it is an objective fact concerning states of affairs which determines whether the lesser evils calculation will come out in favour of the accused. The necessity principle, in contrast, involves a broader assessment which considers the prior circumstances of the accused. In the domestic abuse context, it asks us to consider the efficacy of a person's option to leave their abuser – inevitability is only concerned with the probability of an attack.

In other words, inevitability is not concerned with the reasonable beliefs of the actor; it is only concerned with the probability of harm. There must, of course, be a margin for error when making this probability calculation, such that some 'true' cases may turn out to be false positives if the circumstances had been allowed to unfold uninterrupted. However, this margin is much stricter, and therefore more tolerable, than a bare necessity principle allows.[64] The concept of inevitability now identified, this section expands on the role of inevitability, in terms of its relationship with imminence of threatened harm and other principles such as necessity and proportionality, in the determination of self-defensive acts as justified conduct.

8.4.1 Inevitability and high probabilities

[63] Dsouza recognises such a distinction in the context of formulating a duty to retreat, arguing that a firmer distinction should be recognised between the preconditions for in-principle access to force, and the constraints upon the manner in which that force, if obtained, can be deployed: *op cit* at 741ff. He also recognises that the right to use force is engaged when a threat becomes unavoidable, "and there is no need to wait for the materialisation of the threat to also become immediate" (at 743).
[64] This point is considered in detail below.

Ferzan's theory of the imminence requirement places emphasis on the outward act of aggression from victims as being the kind of threat that triggers the right to self-defence, arriving at this conclusion by asking "at what point is it fair to construe the putative aggressor as posing a threat"?[65] In answering this question she dismisses vague notions of a 'threat' (such as the threat two persons stuck in a cave with limited oxygen pose to one another[66]) as being too broad to use as the basis of self-defence, pointing out that those who have sought to "reject the *status quo* have never been regarded as self-defenders".[67] Even with regards to those who harbour evil intentions, Ferzan states that such people have control over whether they will eventually execute those intentions, with many such intentions being conditional on the occurrence of other events. In short, an intention to harm another is not equivalent to acting to harm another since, until the point of action, one might change one's mind.[68]

Thus, what makes self-defensive conduct justified is that the victim comes within "dangerous proximity" of completing their attack on the accused, characterised by an act of aggression.[69] Aggression, as a tangible act, is crucial to Ferzan because if the right to self-defence is broadened to any person that might *potentially* inflict harm, the distinction between offence and defence is blurred.[70] Indeed, without some standard of aggression to mediate the use of self-defence Ferzan suggests that, in the context of state conflict in international law, two states stockpiling weapons in case of increased tensions would be entitled to attack one another.[71] All of this is to say that what is important for self-defence to Ferzan, and specifically for the imminence requirement, is an *unconditional* act by the victim, about which the victim's intentions can therefore be assumed (irrespective of whether, half-way through the attack, they change their mind): what is important is inevitability. It is this outward act of aggression which operates as an objective claim to the existence of an inevitable attack for which the accused acquires a right to evade with deadly force. For Kaufman, no such debate takes place since any threats which fall short of requiring an immediate response are to be dealt with by the state. What is required for both is an inevitability of harm.[72]

---

[65] Ferzan, *op cit*. at 255. See also Dsouza, *op cit*. at 742.
[66] Ferzan would consider such cases as, if anything, examples of necessity: *ibid* at 256, fn.223.
[67] *Ibid* at 256.
[68] *Ibid*.
[69] *Ibid* at 258.
[70] *Ibid* at 259.
[71] *Ibid* at 260.
[72] Or as close to such an inevitability as is possible in real world circumstances.

8.4.2 The necessary safeguards

The 'imminence as inevitability' theory therefore commits to a two-pronged approach to justifying self-defence – one which considers the probability of attack in contention with the circumstances of the accused. This, I argue, is representative of two key aims: the desire to protect both parties' right to life; and recognition of the law's pivotal role in maintaining civil order through the exclusive use of force. The imminence requirement, operating as the gatekeeper to an accused's 'just-deserts-esque' right to respond[73] to outward acts of aggression, is intended to balance the competing interests of the victim's right to life with the accused's rights to act out of necessity, and is necessary to justify the act. Ferzan and Kaufman's focus on this external act of the victim is a feature which all imminence as inevitability theories necessarily share, irrespective of whether their roots are moral, political, or otherwise. Inevitability in this sense should be understood as encompassing a very high probability – it is the outward act of the victim which increases the risk of injury to the accused to such an extent that they are permitted to retaliate to protect themselves. As aforementioned, inevitability is only concerned with the probability of one outcome – the death or serious injury of the accused in the absence of a response. As it is the high probability of harm occurring which creates such a right or permission to respond, it therefore becomes essential to be able to demonstrate this fact, and a temporal requirement provides a highly effective method.

In contrast, the necessity principle as a broader concern seeks to assess retaliation on a holistic analysis of the accused's circumstances, requiring the consideration of multiple probabilities to determine the more desirable outcome, or 'lesser evil'. The most advantageous solution to an accused may not necessarily correspond with the statistical probability of harm taking place. For example, admitting a defence to a prisoner who kills their sleeping cell mate after the latter has threatened to kill them[74] may be regarded as, objectively, a reasonable and necessary solution under the necessity principle despite the probability of the victim harming the accused at that specific point in time being next to

[73] Ferzan, *op cit*. at 257: "As Michael Walzer argues, '[w]hen we stipulate threatening acts, we are looking not only for indications of intent, but also for rights of response…"

[74] This example is roughly based on the facts of the US case *State v Schroeder*, 261 N.W.2d 759, 760 (Neb, 1978), where the defendant was charged with assault for stabbing his cellmate Riggs. He claimed self-defence on the basis that his requests for a transfer had been ignored, and Riggs had previously indicated that he might assault Schroeder that night when he was asleep.

zero. A singular focus on necessity therefore comes at the cost of the certainty that the wrongful act being avoided would have materialised.

To be clear, the necessity principle is still concerned with the sanctity of life. For example, in the absence of a duty to retreat and a more general responsibility to seek out alternative, reasonable courses of action, a system which focuses solely on a triggering act of aggression risks unacceptably eroding any emphasis on the sanctity of life in favour of archaic notions of defending one's honour,[75] or potentially implying a moral forfeiture of the victim's right to life.[76] Nevertheless, imminence of threatened harm cannot be regarded as just another 'factor' in the necessity calculation, as imminence as necessity theories would suggest. Both necessity and imminence of harm must be considered in the context of a third, proportionality test. The imminence requirement asks us to find commensurability between an unlawful act and a threatened danger; this initial step is necessary to begin any justification calculation. If such commensurability can be established, a proportionality test then asks us to weigh the relative rights of the accused and the victim, in terms of necessity, to determine whether the act committed was truly justified or not.

8.4.3 The rationale of inevitability

The above discussion reveals a broader issue: theories such as Ferzan and Kaufman's only take us so far because while they might do a good job of explaining *how* the imminence requirement is independent from necessity, they fall short when telling us *why*. This is a symptom of failing to engage with the broader underlying principles that inevitability represents. We saw above that while Ferzan correctly identified the importance of an advanced act of aggression, her moral theory did not stand up to scrutiny in terms of explaining why self-defence justified an accused retaliating against an innocent aggressor. Equally, Kaufman's theory can be accused of providing more of an explanation than a justification for an imminence requirement. It is quite clear, on Kaufman's view, that imminence of harm represents the point at which state intervention becomes impossible, and thus citizens become permitted to use force. But absent a broader theory to

---

[75] See, e.g., Fletcher, *Rethinking Criminal Law*, *op cit*. §10.5.4., at 865; PH Robinson, *Criminal Law Defenses* (1984) at §131(d)(3).

[76] For a vocal opponent to this see J Dressler, "Battered Women Who Kill Their Sleeping Tormentors" in A Simester & S Shute (eds), *Criminal Law Theory: Doctrines of the General Part* (OUP, 2002) pp.259-282 at 270-2 and "Battered Women and Sleeping Abusers", *op cit*. at 465-6.

explain *why* citizens are not expected to passively submit to death,[77] Kaufman's theory begins to look suspiciously like an imminence as necessity theory.

Indeed, though he suggests that what matters to the justification of self-defence is the inability of effective state intervention, Kaufman fails to adequately explain on what underlying basis a person becomes vested with the permission to use force. This leads to some difficult questions for Kaufman's theory. For example, without further explanation his theory struggles to explain why a person is not permitted to respond with force to a *lawful* attack, such as a citizen's arrest. If all that matters for the permissibility of such rights is the inability of the state to intervene, then it is unclear why a person should not be able to ward off any attack, legitimate or otherwise. Kaufman argues that the proxy theory is inaccurate because it is easy to conceive of situations involving imminence of harm but not necessity,[78] but if the sole reason a person is entitled to use force is because the state is unable to protect them it becomes hard to see how this amounts to anything other than an acceptance of the imminence as necessity theory.

Thus, by failing to explain the rationale for self-defence beyond pointing to the practical limitations of state involvement, such a political theory appears to undermine its own argument. It is in this sense that Kaufman's theory may be accused of being too naïve – it assumes that in all cases the state will be able to effectively intervene into personal conflicts, right up until there is an imminent attack. But situations like that of Judy Norman,[79] and those of illegal immigrants trafficked into hostage-like situations fearful of deportation, demonstrate quite clearly that this premise is false. At the very least it asks us for a justification which is more robust.

Thus, any explanation for the imminence requirement cannot rely solely on its function as an indicator for the transference of the right to use force, or else it runs the risk of collapsing back into an imminence as necessity theory, specifically a defence where a person responds because they have no other option. We must be able to explain why inevitability of harm is important to situations like self-defence, independent of the necessity principle, which permit the accused to commit a *prima facie* offence. Indeed, stating that imminence of harm is required because it provides evidence of inevitability of

---

[77] Kaufman, *op cit*. at 354.
[78] *Ibid* at 350. Cf. Allhoff, *op cit*. at 1541ff. arguing that self-defence cases supposedly decided on the basis that the imminence requirement was not met could equally be analysed as cases where necessity requirement was not met.
[79] *State v Norman* 324 NC 253, 378 S.E.2d 8 (1989). For commentary on the ineffectiveness of Mrs Norman's alternative options, see Murdoch, *op cit*. at 214-16.

harm naturally raises the question of why a high probability is important, or rather *what* a high probability demonstrates. Our intuitions tell us that a high probability is important because it minimises the scope for a mistaken belief – specifically the belief that the victim intended,[80] and indeed was going to carry out, their attack. For if, with the benefit of hindsight, it is proven that this was not the case, then the accused's act of harm against the victim would be unjustifiable, and the harm experienced by the victim wholly undeserved. This is because, *pace* Fletcher,

> "[B]eliefs alone cannot justify the infliction of violence on another human being. Reasonable beliefs can excuse wrongful aggression against another person but they cannot justify that aggression. Some interaction in the real world is required for a claim of justified harm."[81]

In other words, if a person wishes to justify their conduct, they must be able to demonstrate that the circumstances were as they perceived them to be. It is thus a legal fiction of sorts, where an accused must be able to prove that, had they not successfully thwarted the attack, they would have suffered harm/been killed. In the absence of a crystal ball, imminence of harm can provide proof of the otherwise inevitable harm an accused faced.

8.4.4 Retributivism and probabilities

With the above argument in mind, Yaffe suggests that it is the retributivist principle of just deserts which explains the strict imminence requirement in self-defence, on the basis that a bare necessity principle cannot explain why the victim deserves to be 'punished'.[82] Following just deserts reasoning, if necessity calculations were computed in a vacuum to determine whether conduct was justified, they would run the risk of authorising retaliation against victims in advance for crimes that they may never commit. The paradox emerges, of course, because under a successful prevention such as self-defence, the accused does *not* commit the intended crime. On what basis, then, is self-defensive action justified? Put differently, how can inevitability satisfy our intuitions about preventions being justified when necessity cannot?

---

[80] In the narrow sense of acting with purpose (i.e. so as to include aggressor who lack legal capacity).
[81] Fletcher, "Domination", *op cit.* at 563-4.
[82] G Yaffe, "Prevention and Imminence, Pre-Punishment and Actuality" (2011) 48(4) *San Diego Law Review* 1205.

Yaffe states that just deserts reasoning dictates that it is preferable that a crime be committed because then the harm inflicted is deserved. Yaffe refers to this conclusion as the Actuality Principle – the idea that it is never acceptable to 'punish' individuals for crimes they do not commit.[83] Nevertheless, Yaffe suggests that some preventions can be justified despite this conclusion.[84] Indeed, and generally speaking, all preventions are forms of 'pre-punishment' which, in turn, are breaches of Yaffe's Actuality Principle. However, 'deserved preventions' stipulate that the prevention of a crime, which necessarily involves inflicting 'punishment' on the person who would commit it, *can* be justified by appeal to the Desert Claim; the claim that the harm the aggressor suffers is deserved for the act the aggressor otherwise would have performed.[85] This is true despite the existence of the Actuality principle.[86] Counterfactual considerations about desert are relevant here, says Yaffe, because the justification of prevention through appeal to the Desert Claim happens by engaging in a lesser evils style comparison of the two relevant possible outcomes – one where the crime is committed, and another where it is prevented – in a world in which desert is maximised.[87] In the nonprevention world, the aggressor commits the crime and receives a deserved punishment. Conversely, in the prevention world the aggressor is prevented from committing the crime because of an undeserved 'punishment', but the crime is prevented.

Thus, what is constant in both worlds is that the aggressor receives punishment, but in one scenario a crime is committed, and in another it is prevented. There is therefore a competition between our desire to prevent crime on the one hand, and our pursuit of just punishment on the other. Yaffe suggests that if the calculation is to come out in favour of the prevention world, the degree to which the harm is undeserved should be minimised.[88] This can only be done, says Yaffe, if the aggressor gets as close as possible to performing the criminal act while still being prevented from doing so – the closer the aggressor gets to completing the crime, "the closer the prevention world is to a world in which [the crime] is actualized".[89] In practical terms, this means that the 'prevention world' must be one in which the aggressor's performance of the crime is imminent.[90] This is because to realise a

---

[83] *Ibid* at 1209.
[84] *Ibid* at 1217.
[85] *Ibid* at 1219.
[86] *Ibid*.
[87] *Ibid* at 1220.
[88] *Ibid*.
[89] *Ibid*.
[90] This may, of course, result in a difficult question of deciding what is 'imminent enough' for the desert theory; is it when the aggressor approaches the accused, knife in hand, or must it be when they have made the first swing? The latter is arguably closer to the crime being completed (and thus the 'closest' to the actual

world in which the crime is imminent is to be the closest to a world in which the recipient of the harm is given what they deserve.[91] Where actuality is impossible, we must strive for the 'next-best alternative', and an imminence of harm requirement provides this.[92] Yaffe points out that this conclusion still falls short of giving the victim what is truly deserved, but it does so to an acceptable level in the furtherance of the good outcomes that are realised by preventing the crime.[93]

8.4.5 Killing as a last resort: the sanctity of life

Yaffe's argument suffers from a fatal flaw by characterising self-defensive force as a form of 'punishment'. Many will disagree with this characterisation,[94] and it is particularly perverse in jurisdictions such as Scots law which do not recognise capital punishment. Indeed, it is unclear why anyone other than the state themselves is capable of punishing others,[95] and Yaffe neglects to explain how such a theory is not tantamount to the retroactive approval of vigilante justice by the state. Further, Yaffe's view falls into the same ethical traps of Ferzan's moral theory in that it struggles to explain why aggressors who lack capacity deserve punishment – in no version of the ideal world, which Yaffe claims the criminal justice system is striving for, does the incapax aggressor receive anything close to a death sentence.

Nevertheless, Yaffe's Desert Claim appears to touch on something intuitive which *does* characterise how inevitability is understood in self-defence: it correctly identifies the need to respect both parties' personal autonomy as far as possible in the circumstances. In the context of justified conduct, the rights of the victim cannot simply be unilaterally

---

world sought), but presumably Yaffe would not require the accused to wait until the knife was being thrust at them before they qualified for a right to defend themselves. Presumably Yaffe would state that in such circumstances the competing principles (such as crime prevention) would override such a strict and pedantic interpretation of the imminence requirement, but it is still not easy to answer when and why such principles *would* override.

[91] Yaffe, *op cit*. at 1220.

[92] *Ibid* at 1207.

[93] *Ibid* at 1221.

[94] Fletcher has been explicit in stating that self-defence is precisely *not* a form of punishment: Fletcher, "Domination" *op cit.* at 561-2. See also Bakircioglu, *op cit.* at pp.45-6, 215 & 238. Cf. the US case of *State v Stewart* 243 Kan. 639, 763 P.2d 572 (1988) at 579 where, in rejecting the use of self-defence in a case where a battered spouse killed her abuser in a nonconfrontational setting, they described the attack as a form of "capital punishment" which, if self-defence was permitted, "would in effect allow the execution of the abuser for past or future acts and conduct".

[95] At least in the objective sense. A person can subjectively believe that they are punishing another, irrespective of whether it qualifies as such under the criminal justice system. On the state's exclusive right to punish see, e.g. HLA Hart, *Punishment and Responsibility* (2nd edn, Oxford University Press, 2008) at p.5; Fletcher, "Domination", *op cit.* at 558; Bakircioglu, *op cit*. at p.238.

discarded where the accused can show some pathology or subjectively perceived lack of alternatives. This is because, as identified by Leverick and alluded to above, the law must uphold the sanctity of all life, as far as possible, in its pursuit of justice and civil order.[96] It is precisely because the law must respect each citizen's right to life that it must be so strict with determining when the intentional killing of another was permissible (as opposed to merely excusable).[97]

When such danger is inevitable in the relevant sense, the law permits the accused to prefer their own life[98] to that of the aggressor (subject to a proportionate response), but not before.[99] The imminence requirement, satisfying the inevitability of harm threshold, therefore provides the greatest respect to both parties' right to life and personal integrity.[100] It asks citizens and society to continually respect both lives until we can only choose one, that point being when it is objectively reasonable to assume that serious harm to someone is inevitable. At that point, and only then, we accept that the accused is entitled to prefer their own life. This solution not only has normative value but finds descriptive force in the fact that the right to life is enshrined by Article 2 of the European Convention on Human Rights, and courts adherence to this right is enforced by the Human Rights Act.[101]

There may be other, equally persuasive considerations apart from the sanctity of life which provide justification for preventions despite a lack of desert, such as a focus on crime prevention[102] (which itself feeds into the idea of preferring the accused's life to that

---

[96] See section 8.3.2. On the historical foundations of this principle as championing preservation of life over vengeance and the 'accidental nature' of true self-defence, see C Kennedy, "Defences: Justification, Excuse and Provocation" in M Hill, N Doe, RH Helmholz & J Witte (eds), *Christianity and Criminal Law: An Introduction* (Oxford: Routledge, 2020), pp.253-68 at p.257. See also Bakircioglu, *op cit.* at pp.14, 56 & 209.

[97] I am not alone in this view: see Goosen, *op cit*. at 73 citing the South African case of *S v Makwayane* 1995 6 BCLR 665 (CC) both stating that the right to life is antecedent to all other rights in the constitution. See also Dressler, "Battered Women Who Kill Their Sleeping Tormentors", *op cit*. at 275 and "Battered Women and Sleeping Abusers", *op cit*. at 467; Bakircioglu, *op cit*. at p.45 & 220.

[98] Or, indeed, will be willing to characterise the death as a regrettable accident which occurred while preserving one's life: Kennedy, *op cit*. at p.257.

[99] Cf. Bakircioglu, *op cit*. at p.42, who points out that most jurisdictions with a duty to retreat requirement qualify this such that the notion of 'safety at home' prevails over the sanctity of life.

[100] Gauthier puts it this way: "One central role of the imminence requirement is to ensure that everyone may expect to benefit by maximising the extent to which deterrence and punishment replace pre-emption and retaliation": Gauthier, *op cit*. at 619.

[101] 1998, specifically ss1(1)(a), 1(2), 6(1) & (3).

[102] Yaffe challenges the importance of crime prevention in the context of distinguishing necessity from imminence, suggesting that it cannot be the only justification for self-defence otherwise an accused would be justified to act today to thwart an attack they know will happen tomorrow, and this would be to realise a world "that is further from one in which the person harmed gets what is deserved than the world that results from acting tomorrow": *op cit*. at 1221-2. However, Yaffe's argument is logically incoherent. If, as in Yaffe's proposed hypothetical, I *know* that the crime will be committed tomorrow (i.e. the necessity claim is true), then the desert of the aggressor does not change between now and when the attack is imminent – if what matters to imminence is *demonstrating a sufficient degree of certainty* that an aggressor would have committed the crime, then obtaining that certainty in advance makes imminence redundant, since the ideal and actual worlds will not materially change between now and when the attack is underway. In other

of the victim). In the context of self-defence, however, what is important is that the bar is set very high for citizens to resort to violence against each other. More importantly, utilising sanctity of life as an underlying rationale for the imminence requirement provides a much more satisfying basis for distinguishing it from the necessity principle – it provides an effective counterweight in the proportionality assessment which compares the necessity of the accused's circumstances to the probability of the harm. In this sense, both considerations of necessity and the sanctity of life can be viewed as top level principles which both coexist and compete with one another to provide the rationale for lower-level requirements which make up the framework of justificatory reactive defences.

Leverick has recommended replacing the temporal requirement in self-defence with an 'inevitability of harm' test, comprising two conditions: that the harm threatened would inevitably have occurred (without the accused's intervention); and that there was no reasonable prospect of preventing the harm by taking an alternative course of action.[103] Leverick's test for self-defence therefore incorporates both necessity and inevitability as foundational requirements. If the harm was either too speculative, or there were alternative courses of action which did not involve the commission of a crime, there can be no self-defence. In other words, necessity is a separate consideration from inevitability, which itself is an evolution of the imminence requirement.

Indeed, inevitability in self-defence represents the law's adherence to the sanctity of life principle by placing very strict demands on what kinds of circumstances may justify a violent response. This is to be contrasted with the necessity principle inherent in self-defence which demonstrates the law's acknowledgement of a citizens' competing right, in very narrow circumstances, to utilise self-help such as force when faced with a pressurised situation with no avenue of escape. When the two principles are simultaneously satisfied (the imminence requirement here acting as a safeguard against undeserving claims which may, at a technical level, satisfy the necessity requirement) a person will have a valid claim of self-defence.

---

contexts, the degree of certainty required may not be so high, as where the stakes are much lower, and thus crime prevention may well be a suitable justification in and of itself for the legal prevention.

[103] Leverick, *op cit*. at p.102.

8.5 Temporal Requirements in Other Reactive Defences

This chapter has utilised an analysis of the imminence requirement in self-defence to argue the broader proposition that temporal requirements derive their importance from their ability to demonstrate an inevitability of harm. Such harm, along with the separate necessity requirement, justifies a person's use of force to avoid it. By only permitting the use of force in self-defence when unlawful harm is otherwise inevitable, the law demonstrates its commitment to respecting the sanctity of all human life and bodily integrity, both as a goal in itself and as a means of maintaining civil order. It is a necessary condition before such conduct can be classified as justified, as opposed to merely excused. This understanding of temporal connections appropriately addresses the competing considerations inherent in such situations: it recognises and endorses both the desire of the accused to avoid harm when they have no access to traditional remedies, and the value placed by society on protecting human life generally. What conclusions can be drawn for other reactive defences?

8.5.1 Objective harms and justified conduct

I contend that this understanding of temporal requirements can be applied to reactive defences generally. Specifically, and as a general principle, circumstances must feature an objective inevitability of harm which breaches the sanctity of life principle to *justify* a person's otherwise criminal act. There is clear evidence that the sanctity of life is an equally important factor in the current coercion and necessity defences.[104] However, two immediate challenges emerge which require clarification. The first is that, by committing to a connection between temporal requirements, inevitability and the sanctity of life, this theory necessarily suggests that where harm to life is not at issue, temporal requirements cease to be relevant such that unmeritorious defence claims may be sustained where, for example, property rather than personal integrity is at stake. The second is that this theory pays insufficient consideration to the fact that circumstances of self-defence are unique because the victim and the threatener are the same person.

---

[104] Consider the influential, although not binding, English case of *R v Howe* [1987] AC 417 at 456 where Lord Mackay denied the defence of duress to murder owing to the great value the law affords the protection of life. This was followed by the Scottish case *Collins v HM Advocate* 1991 SCCR 898 at 902 where the trial judge stated, *obiter*, that coercion was not a defence to murder in Scots law "because of the supreme importance that the law affords to the protection of human life". Cf. the unreported case of *HM Advocate v Anderson* (2006) where necessity was admitted as a complete defence to murder, culpable homicide and assault": Ferguson & McDiarmid, *Scots Criminal Law: A Critical Analysis* (2nd edn, 2014), para. 21.4.6.

With respect to the first point, the rule that only threats of death or serious injury can ground reactive defences operates as a safety valve, preventing such unmeritorious cases from arising. In other words, it is only where the accused's personal integrity is at stake that they may break the law to avert the potential harm. In such cases the balance of harm heavily favours the accused such that, provided their response is both proportionate and necessary, we might still feel comfortable justifying their conduct in the absence of an imminent threat. Take, for example, Robinson's classic sinking ship hypothetical discussed above at section 4.2.2, where the crew finds a slow leak but the captain demands they continue on their voyage.[105] Robinson points out the absurdity of requiring the crew to wait until the ship begins to sink before committing mutiny,[106] as at that point they will be too far from shore to do anything about it.[107]

Clearly a temporal link fails to recognise what is important here. Murdoch explains it thus: "it is the worsening of the sailors' position in relation to the shore rather than the temporal proximity of the harm that is the impetus for action".[108] The sailors are thus entitled to prefer their lives rather than risk them to obey the captain's ill-conceived orders, and we would want every crew to be as sensible as the one in the example. Thus, where the balance of harms heavily favours upholding the personal integrity of the accused, their actions may still be justified despite a loose temporal connection. Of course, it is precisely the sanctity of life principle which operates in such cases to push the scales in the accused's favour – whenever the right to life is at risk, property rights must be subordinated to its primacy.

With respect to the second point, that self-defence is unique because the threatener and victim are the same person, it seems incorrect to claim this fact as a normative superiority in all cases. It is not immediately clear, for example, that there is a sufficient normative distinction between someone killing a rival gang member to save themselves from deadly threats by their ringleader, versus killing a young child who is unknowingly brandishing a gun and shooting at people, to warrant a defence in the latter but not the former circumstances. It goes beyond this thesis to examine such a normative claim in any detail, but difficult comparisons like this should serve as evidence that a reliance on the

---

[105] PH Robinson, *Criminal Law Defenses* (1984), §124(f)(1).
[106] For present purposes assume that mutiny means no more than disobeying the captain's orders.
[107] See also the Irish and English cases allowing self-defence to charges of possessing firearms and explosives (respectively) where the term 'imminent attack' was interpreted broadly to include anticipation of further attack based on previous experience: *R v Fegan* [1972] NI 80; *Attorney-General's Reference (No 2 of 1983)* [1984] QB 456.
[108] Murdoch, *op cit*. at 212.

identity of the victim may not always be helpful. In any case, even in the archetypal comparison of killing an innocent third party to evade violence with killing a culpable aggressor, nothing in the proposed theory suggests that where inevitability is present the law must justify killing. It does not follow that because inevitability is necessary for killing to be justified in self-defence it is also sufficient to do so in all reactive defences. There may be other policy reasons for disallowing a defence to murder charges in situations of extreme pressure or individual emergency, even where the accused satisfies the inevitability criteria outlined here. It might make sense to allow a defence in such situations at a theoretical level, but it is certainly not illogical to deny one in view of other important goals, such as deterrence.

8.5.2 Objective harm in situations of individual emergency and extreme pressure

If the above argument is accepted, then the importance of any temporal requirement in reactive defences hinges on their being understood in a justificatory sense. Situations of extreme pressure and individual emergency, as conceived of here, are better understood as excuses, although theoretically speaking a situation of individual emergency *could* be regarded as justificatory, as in situations where objectively the actor does the right thing in the circumstances (think of the sink hole variation of the trolley problem example from section 5.2.5 above). Of course, for such situations to be justificatory, and following the argument of this chapter, it would be required to demonstrate that the conduct in question was truly necessary and proportionate to avoid an inevitable harm of death or serious injury. This is particularly true if necessity is ever to provide a defence to murder. Indeed, the only true distinction between justificatory situations of individual emergency and self-defence becomes the nature of the offence itself. Given the extra-legal, normative foundations of situations of individual emergency, this makes sense.

However, for all other situations of individual emergency, and for all situations of extreme pressure, demanding a strict temporal requirement would be theoretically illogical. The basis for exculpation in situations of extreme pressure is based on the extreme emotions experienced by the accused, brought on by a pressurised situation, and the reasonableness of the actor's response. It would therefore betray the purpose of the defence to state that it applied only where there was an imminent threat. Indeed, such a requirement arguably entrenches the current, mechanistic understanding of emotions by instilling the notion that the defence is only applicable when a person has no time to reason. In most

cases the threat will be imminent, and indeed a temporal requirement might still provide strong evidential value in determining whether the accused could be said to have endured a qualifying pressurised situation,[109] but its presence should neither be necessary nor sufficient to determine reasonableness.

Similarly, in relation to situations of individual emergency not amounting to justified conduct, such cases should be treated as analogous to cases of putative self-defence, those where the accused makes an honest but reasonable mistake as to the facts. Again, the primary question is one of determining the reasonableness of the conduct undertaken, and here a temporal requirement would certainly provide more probative value in determining whether the accused's belief was reasonable, given the shifted focus to the external actions and their intended result, as opposed to the emotional response of the accused, but its presence would not itself be necessary to prove this reasonableness. To this end, we might think of the tests in any new formulation of necessity and coercion as being analogous to the statutory rules on self-defence in the Canadian Criminal Code, where imminence of harm is one of many factors to be considered when determining whether the act was reasonable in the circumstances.[110] One might place a slightly greater emphasis on the objective presence of an imminent threat in necessity (i.e. situations of individual emergency), but in neither test would it be a necessary condition.

One question remains: given that situations of individual emergency captures both permissible and blameless conduct, should this fact be recognised in the form of a distinct set of rules between types of individual emergency? Specifically, should there be a justificatory necessity defence requiring the threat be imminent, and an excusatory defence which focuses more on the reasonableness of the conduct, somewhat analogous to the current English approach? We first must ask to what extent we wish the normative concept to be transplanted completely into the legal framework, assuming this is even possible. Indeed, the existence of competing goals and principles in criminal law suggest that normativity can at best guide reforms, but never dictate them. The rejection of coercion and necessity as a defence to murder serves to illustrate this point. Thus, one might suggest that despite necessity (understood in the context of situations of individual emergency) having the *potential* to include permissible conduct, for practical purposes its requirements should conform to its lowest baseline, which would be blameless conduct. In other words,

[109] Dressler, "Duress" *op cit*. at 271.
[110] See *supra* fn.43.

the rules on necessity should cater to the lesser, reasonable behaviour standard on the understanding that justified conduct will also meet that criterion.

It may be argued that this approach lacks a coherent communicative effect. Insofar as we believe that the criminal justice system should communicate effectively with its citizens, the idea of characterising demonstrably positive acts as merely blameless will not be satisfying. To the extent that we wish the law to speak with precise normative language then, we might prefer to split the defence of necessity in two, allowing for a stricter, justificatory version of lesser evils necessity which adheres to an inevitability requirement, along with an excusatory version of necessity which focuses on whether an act was reasonable in the circumstances.[111] However, the English experience of segregating necessity into several defences may give one pause for thought – it is quite likely that these divisions have hindered rather than aided the communicative effect of the law in this area.[112] I am not convinced that strict adherence to this normative division would generate useful results in practice.

8.6 Conclusion

This chapter has sought to establish the rationale for temporal connections in reactive defences generally, through an analysis of the imminence requirement in self-defence. It has argued that imminence of harm should be understood separately from the necessity principle in self-defence, on the basis that it serves to demonstrate an objective inevitability of harm, which is a separate consideration from the necessity of the situation, and an important factor in justifying the accused's conduct. Both moral and political theories of self-defence point to this concept as providing a crucial role for imminence of harm in the justification calculation. The necessity principle demands that no other option was available, but a close temporal connection separately ensures that the self-defensive use of force was not premature, as a way of ensuring respect for the victim's right to life. If imminence of harm were to be treated as just another factor in the necessity calculation, it could lead to intolerable breaches of the sanctity of life principle, which the law must strive to avoid.

---

[111] Dsouza would presumably support this option, on the basis that such a justificatory form of necessity would not conform to the paradigm form of rationale-based defences on his quality of reasoning hypothesis: *Rationale-Based Defences in Criminal Law* (Hart Publishing: 2019). See, e.g., chapter seven.
[112] See section 3.3.2.

In the context of other reactive defences, such as situations of extreme pressure and individual emergency where their nature is not justificatory and what matters is the overall reasonableness of the response, this translates to the conclusion that, while temporal requirements may provide probative value in determining whether the person was subject to a qualifying situation, their absence should not bar a defence.

9. The Nature of the Harm Threatened

9.1 Introduction

In the previous chapter we began an examination of current necessity and coercion requirements in the context of situations of extreme pressure and individual emergency to determine what role these requirements might play in any new defences scheme. This chapter continues that discussion by examining one further issue in the context of this new framework: the nature and extent of the threatened harm required in reactive defences (hereafter 'threatened harm requirement'). As discussed in chapter three, currently only threats of serious injury or death can ground pleas of coercion or necessity in Scots law – all other types of threat are insufficient. This chapter examines whether the narrow 'only threats of death or serious injury' benchmark is normatively appropriate, or unnecessarily excludes valid categories of cases, such as victims of human trafficking (VOT) who follow the orders of their traffickers while being deprived of their liberty, or those who break the law to alleviate chronic pain. To this end I explore the rationale and significance of limiting what kinds of threat suffice, looking to historical context and how the distinction between a one-off event and on-going ordeal might help explain the purpose of this rule and its broader function. It is submitted that, while the requirement has explanatory value in terms of outlining the kinds of conduct we wish to be included in (or, more accurately, excluded from) reactive defences, the requirement as currently understood is unduly narrow, and that the rule in situations of extreme pressure and individual emergencies should be relaxed to include instances of human trafficking and other kidnapped offenders.

9.2 The Threatened Harm Requirement

As discussed in section 3.4.2 above, to sustain a plea of necessity or coercion under current Scots law, the danger an accused faced must be of death or serious injury. The only exception is that 'serious injury' appears to have been interpreted by the courts to include sexual assault.[1] Other, lesser threats, such as damage to property, are therefore insufficient.

[1] *D v Donnelly* 2009 SLT 467. Although cf. the English case of *R v S(C)* [2012] 1 Cr App R 31, where the court rejected an interpretation of the threatened harm requirement in necessity as covering 'serious injury *or harm*' in relation to allegations of sexual abuse of a child made by her mother against the father in response to charges of child abduction. The court emphasised that only an imminent serious injury could ground the defence.

The presence of such a requirement means that no defence exists where a person damages property to protect their own property, an example which some may argue *should* warrant a defence, owing to the proportionate nature of the harms committed and avoided, and the fact that personal integrity is not at stake. However, this argument can be turned on its head to hold instead that, as only a lesser interest (than personal integrity) is at stake, indeed one which is far easily remedied *ex post* (property can be replaced), one should be prepared to suffer the damage and seek any appropriate redress later, rather than engage in unauthorised self-help.[2] Indeed, there are good policy reasons for holding that a person should not be able to 'deflect' harm directed at their property to the property of others, on the basis that such actions simultaneously erode respect for property rights and the ability of the legal system to provide an adequate response.

Nevertheless, the expectation that persons should be able to withstand all threats other than those of death or serious injury has been challenged in the English courts, with Lord Simon pointing out that:

> "a threat to property may, in certain circumstances, be as potent in overbearing the actor's wish not to perform the prohibited act as a threat of physical harm. For example, the threat may be to burn down his house unless the householder merely keeps watch against interruption while a crime is committed. Or a fugitive from justice may say, 'I have it in my power to make your son bankrupt. You can avoid that merely by driving me to the airport.' Would not many ordinary people yield to such threats…?"[3]

Lord Simon did then acknowledge that as a matter of policy permitting such threats would make the defence (of duress) too broad, and thus concluded that a line should be drawn between threats to property and threats to the person.[4] This seems sensible. While there may be a degree of sentimentality attached to certain property, in many cases we have the infrastructure to effectively replace lost property (i.e. insurance and criminal injuries compensation schemes). It is a shrewd public policy choice to insist that sentimentality alone cannot trump the dictates of the law.

---

[2] Ferguson and McDiarmid give the hypothetical of a person being told by their boss that they are running late for an important meeting, and that if they do not increase their speed (and thus commit the crime of speeding) they will be fired for the proposition that lesser threats ought to suffice for lesser crimes: PR Ferguson & C McDiarmid, *Scots Criminal Law: A Critical Analysis* (2nd edn, Edinburgh University Press, 2014) at p.549, para. 21.3.1. However, this argument fails to account for alternative, legal recourses that such a person could pursue instead, such as a claim for unfair dismissal.

[3] *DPP v Lynch* [1975] AC 653 at 686, per Lord Simon.

[4] *Ibid*.

However, even if we are to accept that in the interests of public policy we should not extend the threatened harm requirement to include property rights, Lord Simon's observation hints at a category of rights where we might find the public interest argument less persuasive. Specifically, it should be noted that death or serious injury does not exhaust the potential ways one might have their right to personal integrity infringed upon. As well as these infringements, and in addition to the relatively underdeveloped exception of sexual assault, a person may be subject to minor injuries or infringements on their personal autonomy, such as the deprivation of their liberty. We might therefore ask on what basis death or serious injury has been chosen as the exclusive forms of threatened harm sufficient for a defence and whether, in fact, other rights of personal integrity might meet this underlying rationale to also be considered legitimate threats for the purposes of situations of extreme pressure and individual emergency.

In relation to minor injuries, issues of proportionality point away from a defence in the context of serious crimes such as arson or robbery: in such cases we would expect persons to endure the injury in the interest of upholding the law and preventing an arguably greater societal harm. However, such scenarios are both unlikely and conceptually easy to solve; both more likely and difficult in practice are those where the harms are proportionate in the sense that both the harm threatened and harm committed to avoid it are relatively minor. However, and again, for public policy reasons we might expect persons to endure such harm to uphold the greater interest of maintaining civil order, rather than to 'pass on the harm'.[5] While we might have sympathy for those who choose to prefer their own interests in such situations, this sympathy can only extend to mitigation in sentencing rather than full exoneration, for such harm can never outweigh the commission of a crime.[6]

In contrast, attacks on personal autonomy present a more complex issue because the wrong envisaged by such attacks is grounded in the rejection of a person's right to self-determination through subjugation, rather than any physical injury per se, and thus such attacks can be characterised as the infliction of a form of psychological harm.[7] It is in this sense that a relatively minor physical interference can become an appropriate trigger for

[5] Although this might not always be clear. See, e.g. *Pipe v DPP* [2012] EWHC 1821 (Admin) where it was argued that a defence of necessity should have been open to a man who broke the speed limit taking a child with a broken leg to the hospital. The appeal was granted and the case remitted to the Magistrates' court, but there appears to have been resistance to the remittal by defence counsel on the basis of a prior dispute with the magistrates about the seriousness of the injury and whether it amounted to a defence.

[6] In this sense, we might draw an analogy to provocation in assault cases, which is available only as a mitigating factor.

[7] The subjugation of a person may involve force, but it is not an essential feature of the harm caused/wrong experienced by sexual assault or the deprivation of liberty.

reactive defences, as where touching is sexual in nature. Equally, there may be no physical interference at all, as where a person is locked in a room against their will. Intuitively, we can understand why a person would wish to avoid psychological harm and might break the law to avoid it. Indeed, to some extent we can say that *all* threats have a psychological component to varying degrees – this is particularly true where the threatener proposes to physically injure someone else close to the accused – and this fact has helped form the understanding of situations of extreme pressure presented here. Furthermore, in the context of criminal offences there is precedent for defining harm as including psychological injury,[8] and thus it is not a leap to suggest that our understanding of threatened harms in reactive defences *should* accommodate the harm associated with deprivation of liberty.

One final aspect to consider is where the harm experienced by the accused is physical, and indeed extremely severe, but *internal* to the accused, as where a person suffers from a chronic, debilitating condition. Here, consider *Attorney General's Reference (No 2 of 2004)*,[9] where the accused had all been charged with the cultivation of cannabis and claimed necessity on the basis that cannabis was the only drug capable of alleviating their chronic pain (without equally debilitating side-effects caused by other forms of legal medication). Questions arose about the intended meaning of 'injury' and whether it is broad enough to accommodate the concept of pain. The appeal court rejected such a broad conception of injury to include conditions the accused was already suffering from as too subjective,[10] while reinforcing the view taken in *R v Rodger and Rose*[11] that only threats that are extraneous to the accused could ground a defence.

However, an argument can be put forward that chronic pain should be considered analogous with psychological harm, on the basis that the latter is commonly understood as a type of pain experienced. If the torment of being sexually assaulted or watching a loved one be hurt is a sufficient threat, it is not immediately clear why severe pain from an old injury or being held captive are not. It is at least debatable whether threats in the former category are, in all cases, any more or less harmful than the latter. There is thus an incentive to reassess the term 'serious injury' in the context of the threatened harm requirement, to determine its precise limits and what they ought to be for any framework proposed here. It is clear that being falsely imprisoned or experiencing chronic pain are not trivial matters, and thus cannot be regarded as part of a *de minimis* exception of harms that

[8] *R v Ireland; R v Burstow* [1998] AC 147.
[9] *R v Quayle; R v Wales; R v Taylor and another; R v Kenny* [2005] 2 Cr App R 34.
[10] *Ibid* at para [78].
[11] [1998] 1 Cr App R 143. Discussed at section 3.4.3 above.

persons are expected to bear, as with minor injuries. Is it reasonable to reject breaches of personal autonomy and chronic pain as appropriate threatened harms based on objectivity concerns, or are there other factors at play? The rest of this chapter shall consider these questions.

9.3 The Impact of Historical Development

The topic of whether subjugation and chronic pain are appropriate harms for the threatened harm requirement did not arise until relatively recently. Indeed, sexual assault was clearly not envisaged in the original formulation of these defences espoused by Hume,[12] although it now forms part of the requirement. This is more likely due to an absence of cases rather than any implicit rejection of the concept. Nevertheless, the historical context may still prove to be illuminating here. Glazebrook points out that the term 'Homicide by necessity' was given first to killings in self-defence in most English criminal law treatises between the thirteenth and nineteenth centuries,[13] and thus it is possible that the threatened harm requirement in necessity (and thereafter, by extension, coercion/duress) was influenced by this early connection between the two defences.[14] In other words, it is possible that the jurisprudence surrounding the more commonly pled defence of self-defence played an influential role in determining what threats of harm could ground pleas of the modern necessity (and coercion) defence(s). Indeed, we also saw in chapter three that there was an extended period of time where neither coercion nor necessity were fully recognised in Scots law until the appeal court decisions of *Thomson v HM Advocate* in 1983[15] and *Moss v Howdle* in 1997[16] respectively.

More importantly, when necessity was finally recognised as a substantive defence in 1997, it was on the basis that it was analogous with coercion, and therefore operated on the same underlying basis found in Hume.[17] This is an important factor, because Hume was discussing a very narrow conception of compulsion – one which focused on the liability of persons caught up in a rebellion, who could not be expected to resist the force of a violent gang. In other words, Hume's early conception of what would become coercion was very much tailored to accommodating alternatives to self-defence: only if the force against

---

[12] For more detail see section 2.3.
[13] PR Glazebrook, "The Necessity Plea in English Criminal Law" (1972) 30(1) *Cambridge Law Journal* 87 at 110.
[14] For the historical connection between Scots and English law, see section 2.4.2.
[15] 1983 JC 69.
[16] 1997 JC 123. See section 3.1.
[17] See section 3.2.1. See also the comments by Lord Carloway in *Van Phan v HM Advocate* [2018] HCJAC 7 at [43] where he refers to self-defence as "related" to coercion.

one's person was insurmountable could that person concede to the threat for the duration that such a force exerted itself on their will.[18] Thus, insofar as we can point to historical markers to determine the rationale for the requirement, a historical account places the focus of the threatened harm requirement on an immediate need for action to protect one's life. The later inclusion of sexual assault places the importance of this second aspect, protection of life, in doubt, but the idea of an immediate need for action can be seen to bar the way to including threats such as deprivation of liberty or chronic pain within the scope of the threatened harm requirement.

This descriptive account of the close relationship between coercion/necessity and self-defence in their historical evolution might also explain the seemingly arbitrary line drawn between the types of psychological harm which can satisfy the threatened harm requirement. If, as is argued here, the defences of coercion and necessity evolved from a general self-defence mould, it would explain why a person is permitted to respond to threats aimed at harming others, but is not presently permitted to respond to other threats which do not pose physical harm to the accused. Neither subjugation nor chronic pain represent an immediate attack on the accused or anyone else, and thus do not fit the characterisation developed for situations that engage reactive defences.

9.4 Issues of Characterisation

Chapter eight examined the imminence requirement in self-defence and suggested that the importance of such requirements rests in their ability to demonstrate an inevitability of harm, something which is crucial to prove before the state can permit the taking of human life. To this end, Robinson tells us that a person taken prisoner and told by their captor that they will be killed at the end of the week should not have to wait for the threat to materialise to protect themselves from harm if an opportunity arises before then.[19]

---

[18] We see similar language emerging in the Scottish courts on this basis: e.g. *Cochrane v HM Advocate* 2001 SCCR 655 at 661 requiring that the accused should expect to be 'attacked immediately'. Also consider the recently abolished 'defence' of marital coercion in English law which did not require a threat of harm from the husband – only that the accused could be said to be under the control of her husband, often taken to be assumed by his presence at the crime: *R v Shortland* [1996] 1 Crim App R 116 at 117-8. The law has therefore, at least in some spheres of society, previously accepted that one can be coerced/overwhelmed without threats of violence. This analogy is, however, of limited strength given the confusing and complex nature of this defence/presumption, including issues with its characterisation as a defence at all. See, e.g.: G Rubin, "Pre-dating Vicky Pryce: The Peel Case (1922) and the Origins of the Marital Coercion Statutory Defence" (2014) 34 *Legal Studies* 631; E Ireland, "Rebutting the Presumption: Rethinking the Common Law Principle of Marital Coercion in Eighteenth- and Nineteenth-Century England" (2019) 40 *Journal of Legal History* 21, particularly at 27ff.

[19] PH Robinson, *Criminal Law Defenses* (1984), §131(c)(1), p.78.

However, presumably our opinion about Robinson's hypothetical captive's right to employ (lethal) force is different if the captor never makes such an announcement. If, instead, the captive is made to carry out unlawful activities and threatened with violence if they refuse, we might have greater pause for thought about endorsing this person's use of lethal force to escape.[20] They can, after all, continue to carry out the unlawful activities, thereby never experiencing any physical harm. In the context of criminal charges for the commission of those unlawful activities, however, or if the captive otherwise broke the law in the course of making their escape, we might feel more sympathetic about their claim that they carried out such actions because they were held against their will.

The fact that the threatened harm requirement is identical across reactive defences in Scots law contributes to a broader issue of characterisation, whereby it is difficult to differentiate between a set of similar defences where the circumstances captured by each overlap.[21] Indeed, understanding the precise contours of necessity and coercion has been a primary task of this thesis. All reactive defences focus on an immediate response to deadly violence, with there being no reasonable alternative course of action open to avoid the threatened harm. The prescription of identical requirements necessarily results in an increased focus on the underlying normative claim of each defence as a means of defining their individual parameters, with any subsequent failure to understand or engage with these normative claims leading to regression and thus confusion.

For an example of this regression, take the 2012 changes to self-defence in Canadian law, found in section 34 of the Canadian Criminal Code,[22] where it now appears that self-defence has subsumed any necessity claims involving human threats.[23] Section 34 states that, subject to issues of purpose and reasonableness, a person will not be guilty "of an offence" if they believe that force is being used against them, or a threat of force is being made against them. The provision only focuses on the *purpose* of the act in protecting oneself from the use of force – it does not specify what *type* of act an accused must commit – and thus presumably the theft of a car to escape violence, for example,

---

[20] This certainly appears to have been the view of the appeal court in England: *Nicklinson v Ministry of Justice* [2014] 2 All ER 32 at para [54]: "There is no self-evident reason why [the sanctity of life] should give way to the values of autonomy or dignity".

[21] It is precisely because many of the requirements are identical that Clarkson suggested replacing self-defence, necessity and coercion with one defence: CMV Clarkson, "Necessary Action: A New Defence" [2004] Crim LR 81.

[22] RSC 1985, ch C-46, s34, entitled "Defence of Person", amended by The Citizen's Arrest and Self-Defence Act 2012, SC 2012, C-9.

[23] Necessity in Canadian law would appear to still operate in the case of physical dangers which do not involve human interference.

would qualify as self-defence under the provision.[24] Alternative ways of defining defensive conduct from other kinds of reactive/emotional responses become impossible because self-defence and other reactive defences, like necessity and coercion/duress, feature identical, restrictive harm requirements – they all require that the threat to the accused be of death or serious injury.

Further, we saw above in chapter three that the courts have previously determined that the distinction between coercion and necessity is one without a relevant difference,[25] and this conclusion has influenced the version of necessity which was received into Scots law.[26] When the requirements of reactive defences are identical it becomes very difficult, beyond some superficial factual differences (such as the origin of the threat), to characterise and individuate them. This has implications for any normative theory of reactive defences. Horder has attempted to distinguish reactive defences in English law by suggesting that each defence involves a distinct, key moral issue.[27] In necessity cases, this is the existence of a moral imperative to act, even if this might involve wrongdoing, to negate or avoid some other evil. In coercion/duress cases, this is the personal sacrifice the accused is being asked to make, and whether they should be expected to make that sacrifice. Finally, in self-defence it is whether the accused has a permission to act; to take reasonable and proportionate steps to negate or avoid an unjust threat.

One can take issue with Horder's framework on the basis that it does not sufficiently prevent the identified problem of overlap, thereby defeating its purpose. This is because there are undoubtedly situations of duress where we might speak of the accused as having a permission to act, or situations of self-defence where there is an imperative to act.[28] Of importance for our purposes, however, is the idea that such normative distinctions cannot be maintained in a legal environment which stubbornly adheres to a universal

---

[24] Presumably the drafters were concerned with the kinds of examples presented by Horder: "If the only way I can stop a would-be attacker killing me is to release a poisonous gas… then I am entitled to have such a step considered as potentially necessary and proportionate, even though it does not involve the use of force": J Horder, "Self-defence, Necessity and Duress: Understanding the Relationship" (1998) 11 *Canadian Journal of Law and Jurisprudence* 143 at 144.

[25] Section 3.2.1, citing Lord Hailsham in *R v Howe* [1987] AC 417 at 429.

[26] The defence of necessity in Scots law is arguably the English 'duress of circumstances' by a different name: see *ibid*.

[27] Horder, *op cit*. at 143.

[28] Conversely, consider Horder's example of threat 'triggers', where a person is linked up to a door such that the next person to touch the door handle will electrocute them. Horder argues that what distinguishes attempts to prevent the second person from touching the door handle as either self-defensive or necessitous depends on whether the second person's actions are 'free, deliberate and informed', i.e. meant to harm. Horder does not make it clear, however, why knowledge and intention by the person who 'triggers' the harm should dictate how we understand the first person's response to that harm. Why is it clear that the first person has, and only has, a moral imperative to act if the second person is unaware of the implications of their action? Horder, *op cit*. at 153ff.

threatened harm requirement as Scots law does. This is because the legal prescriptions necessarily dictate the parameters of any underlying normative claim each defence purports to represent.

Taking Horder's framework as an example of this point, the 'moral imperative to act' which Horder claims typifies necessity must be understood in the context of the threatened harm requirement, such that anyone acting for reasons other than impending serious injury or death cannot be said to operate under any such imperative. How are we to differentiate a moral imperative from a permission in self-defence, however, when both the permission and imperative are only engaged on equivalent terms?[29] Likewise, how can the sacrifices made under duress to be considered 'personal' when they are described in equivalent terms to permissions and moral imperatives? Horder suggests that necessity involving a 'strong justification' providing an overriding reason to act, in contrast to the 'weak justification' or permissibility inherent in self-defence claims,[30] but a unified threatened harm requirement ensures strong justificatory reasoning in all cases. If one can only act out of necessity or duress when life is at risk, and can only employ self-defensive force when life is at risk, then the reasons for acting must be strong in all cases. Equally, the rigidity of the threatened harm requirement makes Horder's example of a solider killing their badly wounded comrade to 'save them' from being tortured by the enemy an inaccurate example of what would suffice as a moral imperative under anything other than a value theory deriving from the dictates of rationality and categoricity.[31]

A threatened harm requirement can therefore be seen as potentially stifling, depending on how it is defined across reactive defences. Its current construction under Scots law undermines any normative claims the law attempts to make. In addition, and depending on the construction of a jurisdiction's reactive defences framework, a uniform threatened harm requirement may even contribute to the promotion of some defences at the expense of others, such as self-defence becoming the defence of choice for explaining behaviour because its requirements are more favourable to the accused and its scope is broad enough to encompass factual narratives which we would otherwise describe as necessity or coercion. However, even if we decide that characterisation and fair labelling are not important and introduce a nebulous 'reactive defence' to replace the current defences framework, that defence would still be at risk of being under-inclusive in terms of

[29] Note this criticism is not true of English law where 'self-defence' extends to defence of property: see s76 Criminal Justice and Immigration Act 2008.
[30] *Ibid* at 155.
[31] *Ibid* at 155-6.

the types of threats serious enough to warrant exculpation for crimes committed to avoid them. The rest of this chapter shall elaborate on this claim.

9.5 False Imprisonment as a Site for Injury: The Limits of Personal Autonomy

The issues identified above necessarily raises the further question of how the threatened harm requirement should be understood in the context of situations of extreme pressure and individual emergency, particularly given the recent increase of human trafficking cases reaching the Scottish courts where the VOT is charged with offences committed while under their captor's control.[32] If the harm requirement in necessity and coercion has indeed been unduly influenced by their historical association with self-defence and its focus on lethal force, there is *prima facie* no compelling reason to restrict the threatened harm requirement to threats of death or serious injury in circumstances which do not involve the actor killing another. At the very least, there is an argument to be considered that deprivation of liberty, as an important aspect of personal autonomy and certainly not a trivial matter, *should* be included within the scope of the harm requirement.

To this end, *Van Phan v HM Advocate*[33] represents an interesting development in this area. *Van Phan* is an archetypical case of human trafficking for male victims – Van Phan was locked in a tenement flat in Glasgow and told to cultivate cannabis after being promised work in the UK by his traffickers back in Vietnam. Van Phan's status as a VOT was contentious, with some evidence suggesting that he was free to come and go from the property (police found a key to the flat hanging in the only occupied room). For this reason, prosecutors decided to proceed with charges of producing and supplying cannabis against him. His counsel argued that he had insufficient recourse to adequate defences, inconsistent with the Warsaw Convention which seeks to protect VOT from prosecution for crimes committed during their ordeal.[34] They highlighted the absence of any bespoke statutory defence to offences committed while under a trafficker's control,[35] and that the

---

[32] See, e.g., *Van Phan v HM Advocate* [2018] HCJAC 7; *Hai Van Le v HM Advocate* [2019] HCJAC 44. According to Scottish government statistics, the number of persons being referred to the National Referral Mechanism has been increasing every year since the Trafficking and Exploitation Strategy was introduced in 2017, with an increase of 125% between 2018 (228) and 2019 (512). Annual progress reports can be found at https://www.gov.scot/policies/human-trafficking/.

[33] [2018] HCJAC 7.

[34] The Warsaw Convention was adopted into EU law by Directive 2011/36/EU. See specifically recital 14 and art.26.

[35] Such a defence exists in English law under the Modern Slavery Act 2015 s45(1). In contrast, under the Human Trafficking and Exploitation (Scotland) Act 2015 s8 the Lord Advocate must publish guidelines for Scottish prosecutors to help them decide when to bring proceedings against an accused who commits an offence because of compulsion which appears to be directly attributable to their being a VOT. There is no

common law defence of coercion would be inapplicable because of the lack of an imminent threat of death or serious injury, as per *Thomson*.[36] On this basis, the sheriff was unable to reach a decision on the issue of punishment and thus remitted the case to the High Court for a decision.

Lord Carloway, delivering the judgment of the appeal court, disagreed with this assessment of coercion, stating that the defence could nevertheless be pled if circumstances demonstrated than an accused had been coerced into carrying out the act under a threat of violence. He distinguished *Thomson* on the basis that the circumstances there were very different, involving a robbery using firearms. He thus emphasised the fact that *Thomson* concerned a "single violent act, committed in a place accessible through public spaces where resort to the forces of law and order will normally be possible".[37] In such cases there would need to be a threat of immediate and serious violence if coercion were to be made out. In contrast, and citing the requirements for the defence stated in *Thomson* as originally laid down by Hume, Lord Carloway placed emphasis on the specific wording used by the court in relation to the four part test for coercion to argue that, rather than being a hard and fast rule, the imminence of danger was conditional, to be construed within the context of the "facts and circumstances" when determining whether or not a person was truly coerced by virtue of a "genuinely anticipated and unavoidable violence".[38]

Thus, Lord Carloway suggested that if Van Phan had reasonable grounds to believe that by refusing to cooperate he would have been seriously injured by his captors on their return to the property, the defence would be made out.[39] In justifying this conclusion, his Lordship focused on the importance of the proportionality requirement in coercion cases, pointing out that the judgment in *Thomson* correctly stressed the need to strike a fine balance between "the nature of the danger threatened and the seriousness of the crime".[40] Lord Carloway appeared to qualify this justification by agreeing with Lord Wheatley's statement from *Thomson*, that, "Nevertheless, there may be circumstances in which a person is exposed to a threat of violence to himself or a third party from which he cannot be protected by the forces of law and order and which he is not in a position to resist".[41] To

---

equivalent defence if prosecutors do bring charges. These guidelines are currently in equivalent terms to the s45 defence in England, see: *Lord Advocate's Instructions for Prosecutors when considering Prosecution of Victims of Human Trafficking and Exploitation*. These guidelines can be found available at https://www.copfs.gov.uk/publications/prosecution-policy-and-guidance?showall=&start=4.

[36] 1983 JC 69.

[37] *Van Phan* [2018] HCJAC 7 at [42].

[38] *Ibid*.

[39] *Ibid* at [43].

[40] *Ibid*, citing Lord Wheatley in *Thomson v HM Advocate* 1983 JC 69 at 78.

[41] *Ibid*.

this he added that such situations may arise in human trafficking cases; where they do, the effect in law would have to be assessed on its particular facts.[42]

There are several difficulties with Lord Carloway's judgment regarding his interpretation of the relationships between the different requirements in coercion which demonstrate a misinterpretation of *Thomson*, either done accidentally or deliberately with a view to planting the seeds for eventual judicial development in these unique situations. Specifically, Lord Carloway suggests that *Thomson* identified a broad proportionality requirement in coercion, "as in the related defence of self-defence", based around the threatened harm and the offence committed.[43] However, the actual passage from *Thomson* is taken from a discussion about whether coercion could be said to apply to all crimes. Lord Justice-Clerk Wheatley was discussing the meaning of Hume's reference to 'atrocious crimes' and emphasising that due care should be taken when applying this (then novel) defence because, if applied too generously, it may lead to undesirable effects on "the proper and fair administration of justice".[44] Indeed, Lord Carloway's judgment neglects that in the subsequent jurisprudence of the appeal court, Lord Justice-General Rodger disapproved of these comments, rejecting any substantive implications of drawing a line between atrocious and non-atrocious crimes, stating that "[i]t would be an odd legal system indeed which… allowed coercion to elide guilt of the crime of armed robbery, but not guilt of the offence of exceeding the speed limit".[45]

Thus, the court's comments in *Thomson* should be regarded, at best, as merely a warning about coercion's potential for misuse, and Lord Carloway's suggestion that a proportionality requirement had already been established in the defence should be questioned.[46] Lord Carloway appears to make this connection with a view to eroding the primacy of the close temporal requirement; he is in some ways forced to do this because of how strict the threatened harm requirement is, it being stated in unequivocal terms by both Hume and subsequent courts interpreting it. In contrast, the natural scope for interpretation of 'immediacy', along with the subsequent jurisprudence of the appeal court, allows him

---

[42] [2018] HCJAC 7 at [43].
[43] *Ibid*.
[44] *Thomson v HM Advocate* 1983 JC 69 at 78.
[45] *Moss v Howdle* 1997 JC 123 at 126.
[46] He is not the first, however, to suggest this requirement (despite the stringency of the threatened harm requirement). See, e.g., *Lord Advocate's Reference (No 1 of 2000)* 2001 JC 143 at 160, [47]: "As a matter of general principle it appears clear that the conduct carried out must be broadly proportional to the risk. That will always be a question of fact to be determined in the circumstances of the particular case." Cf. *Attorney General's Reference (No 2 of 2004)* [2005] 2 Cr App R 34 at para [46]; and see the statement by Lord Justice-General Rodger in *Cochrane v HM Advocate* 2001 SCCR 655 at 661 which questions the efficacy of a reasonableness test where the threatened harm requirement is restricted to death or serious injury.

some leeway to suggest that, in special circumstances, the requirement of immediacy might be circumvented. It is, nevertheless, not a neat solution. Reliance on a normative assessment of the harm inflicted versus the harm avoided to determine the value of the temporal requirement in any given case results in there being an invisible line where a crime becomes so atrocious that immediate danger is a necessary condition for the defence; and all we know is that armed robbery crosses that line, whereas cultivating cannabis potentially does not.

Indeed, what is most peculiar about Lord Carloway's comments on proportionality here is that it was entirely unnecessary if his purpose was to open the door for a coercion defence in circumstances which lacked an immediate danger. This is because he also relies on the subsequent comments in *Thomson* which state that special circumstances may exist to circumvent the ordinary rules. It is unclear why this statement alone was insufficient to carve out a path for coercion to be pled in trafficking situations where the threats of injury were not immediate in the traditional sense. In any case, the judgment in *Van Phan* presents us with a strong example of where a change to the threatened harm requirement could give a more conceptually pleasing result. If the requirement were modified to allow a defence where an accused is falsely imprisoned, courts and juries could make honest decisions about whether the offences committed were excusable in the circumstances. In effect, it could build into the process the kind of proportionality that Lord Carloway seems to be striving for, without confusing the purpose of reactive defences or their requirements.

There appears to be a broader issue at play here when we consider the relationship between immediate danger and threatened harm requirements in reactive defences. *Van Phan* presents itself as a case concerning the relevance of *immediacy* of harm; I have argued that a solution may be found by examining instead the *nature* of the harm. However, stripping the issue down to its roots, it seems that we should rather be asking more generally what limits can be placed on the concept of 'reacting'. If a person lunges at me with a knife, any act I then undertake in response is clearly a reaction. Equally, however, if a person deprives me of my liberty and forces me to work under intolerable conditions, we may likewise say that my compliance or defiance is a reaction to that conduct. What distinguishes these situations is, as Lord Carloway alludes to in his judgment, the fact that the former scenario involves reacting to an event, whereas the latter involves reacting to an on-going situation (which was created by a prior event). Physical injuries can only be sustained once, and then the action (of being injured) is complete. In contrast, being deprived of one's liberty is a continuous state, such that the 'conduct' to

which the accused reacts exists in perpetuity, at least until something happens to stop it. Discussions of immediate danger are therefore unhelpful. The question thus becomes whether reactive defences should be broad enough to include reactions to actively ongoing conduct.

9.6 The Limits of Psychological Harm and Chronic Pain

To help us answer the question of whether necessity and/or coercion should include reactions to ongoing conduct, we can consider a hypothetical where a person is subjected to an ongoing physical assault: specifically, a person being slapped on an hourly basis. In this scenario, if the accused were to eventually snap and react in a way which breaks the law, it seems inappropriate to evaluate the validity of their reaction by reference to any individual slap (which, under current rules on coercion/necessity, would not meet the strict threatened harm requirement's threshold). We can only really understand their situation by taking account of the cumulative effect of the slaps (which may amount to some combination of both physical and psychological injury). With this logic in mind, it becomes easier to then understand why a person falsely imprisoned might steal a car to escape or submit to farming cannabis, if their alternative is to be repeatedly beaten, even if each beating has a very low chance of seriously injuring or killing them.

However, Lord Carloway's example of an imprisoned person expecting a beating on their captor's return appears problematic for two reasons. First, by specifying threats of death or serious injury, his example fails to capture this broader idea of cumulative effect explored above which is unique to these kinds of situation; his is not a typical example, likely only representing a small minority of cases. Second, it does nothing to challenge the conventional understanding of threatened harms in these situations, despite Lord Carloway's claims that the defence of coercion is broad enough to encapsulate both traditional 'single violent acts' and 'genuinely anticipated and unavoidable violence'. This is because Lord Carloway's understanding of trafficking situations is one where the 'genuinely anticipated and unavoidable violence' sufficient for a defence just *is* a 'single violent act'. On this kind of interpretation, the harm in trafficking is erroneously understood as anticipation of (an appropriate degree of) violence. This undermines much of the trafficking experience.

Indeed, Ormerod has similarly considered that in many cases what makes a threat of false imprisonment effective at overbearing a person's will is the fact that such threats

are intrinsically linked with an implied threat of violence.[47] In such cases he presumes that an accused would be able to rely on a defence on this basis. While this is undoubtedly true, it is important to recognise deprivation of liberty as a separate harm, particularly in the human trafficking space, to adequately reflect the lived experiences of those who are subjected to it. Human trafficking cases involve prolonged abuse which can span a long period of time and engage in a different kind of psychology in terms of the impact on the accused,[48] the dangers they may face, and what we can expect from them.[49] Indeed, in addition to a fear of habitual low-level violence, VOT may be wary of involving state authorities for fear of being arrested and/or deported. In addition, they may experience extreme manipulation with their freedom being used as a bargaining chip for servitude.

A reliance on serious injury or death as the qualifying trigger point therefore unfairly sets the bar too high, undermining these concerns. The experience of VOT suggests that some combination of these fears and manipulation may influence their decision, rather than there being a monolithic fear of death. One can also draw an analogy here with sexual assault – it would be absurd to suggest that fear of physical injury was the primary reason for allowing a reactive defence in such situations: there is clearly something else, stemming from personal autonomy, which the law recognises is unacceptably infringed in these cases. It is fundamental to our liberal democracy that persons should enjoy freedom from any unwanted interference beyond the narrow state interventions necessary for the maintenance of civil order. We saw in the last chapter that, as part of respecting this freedom from interference, persons are permitted to kill others when their life is at stake. Another part of respecting this freedom means acknowledging that freedom itself forms part of the personal integrity the law seeks to protect, particularly when lesser crimes have been committed. An expectation of serious violence therefore cannot be the (only) benchmark if we wish to take coercion in trafficking cases seriously.

---

[47] D Ormerod, "Duress: *R v van Dao (Vinh)*" (case commentary) [2013] Crim LR 234 at 236.

[48] See, for example, the English case of *R v GS* [2018] EWCA Crim 1824 at [46] where it was suggested that the accused's mental state indicated that she was 'vulnerable to exploitation and less able to resist pressure'. Simpson argues that a lack of adequate medical evidence leaves courts "detached from the true nature of modern slavery and the traits and characteristics commonly found amongst victims which can influence their motivations for offending": B Simpson, "Modern Slavery and Prosecutorial Discretion: When is it in the Public Interest to Prosecute Victims of Trafficking?" (2019) 83(1) *J. Crim. L.* 14 at 17.

[49] Simpson, *ibid* at 18: "The reality of their situation is often incomprehensible to individuals whose life experiences are far from the turmoil of trafficking". The standard of fortitude required by VOT in the s45 MSA 2015 defence in England has been a point of criticism, with parallels being drawn to the difficulties of applying a reasonable person test in cases featuring "learned helplessness". See, e.g.: K Laird, "Evaluating the Relationship between Section 45 of the Modern Slavery Act 2015 and the Defence of Duress: An Opportunity Missed?" [2016] Crim LR 395 at 399-401.

Ormerod has considered that the inclusion of deprivation of liberty to the threatened harm requirement would represent a "significant step towards diluting the current rigorous threshold",[50] on the basis that false imprisonments are much more likely to vary widely in terms of assessing the gravity of the threat than are serious injuries.[51] To that end, Ormerod views any expansion of the threatened harm requirement as reducing clarity to juries making their decision. Without empirical data to support this assertion, however, this seems like a weak argument.[52] Indeed, it is not self-evident that the category of 'serious injury' itself is clear for juries. This chapter has demonstrated that 'serious injury' is far more nebulous than it initially appears. Is a broken limb a serious injury? Does the bone break have to be complicated, for example, such that it will leave a lasting injury, or will a clean break with a normal recovery suffice? Are torturing techniques such as nail pulling serious injuries? To what extent is psychological pain a 'serious injury' for the purposes of the requirement? Likewise for sexual assault?[53] These are all equally difficult questions for any jury to answer when deciding if a person acted from a valid threat. The suggestion that the inclusion of false imprisonment would make the assessment intolerably less clear is thus unconvincing. One must also bear in mind that any relaxation of the threatened harm requirement will necessarily be tempered by the fact that other strict requirements still exist in respect of these defences.

In the context of this thesis, false imprisonment appears to be a good example of a situation of extreme pressure – one where the accused is unjustly placed in intolerable circumstances (deprivation of liberty) eliciting a strong emotional response, and we accept that in such a context their criminal actions were therefore understandable. Coercion *qua* a situation of extreme pressure can better accommodate this kind of claim than the current defences framework because it better reflects our intuitions about why VOT should not be treated as blameworthy. The accused does not try to bring about some greater good or right with their actions, so they do not qualify for a situation of individual emergency (i.e. necessity), but they do act based on an emotional pressure which would significantly influence a normal person. Given that sexual assault has already been recognised by the Scottish courts as an adequate threatened harm, and that the Scottish government has made a commitment, enshrined in statute,[54] to protecting VOT from further trauma through prosecution and punishment, it seems sensible to extend the threatened harm requirement

---

[50] Ormerod, *op cit*. at 237.
[51] *Ibid*.
[52] A fact which Ormerod is aware of: *ibid*.
[53] Compare and contrast *D v Donnelly* 2009 SLT 467 with *R v S(C)* [2012] 1 Cr App R 31: *supra* fn. 1.
[54] Human Trafficking and Exploitation (Scotland) Act 2015, particularly at s8.

of coercion to include deprivation of liberty as a relevant harm where it is made out on the facts.

Ormerod is, however, correct to raise concerns about dilution of the threatened harm requirement, as it is a delicate balance. If the pendulum swings too far in the opposite direction, we risk authorising a defence for emotional reactions to a wide variety of threats which, as noted at the start of this chapter, should not be given recognition by the law for good reasons, both in terms of logic and public policy.[55] To this end, it seems that chronic pain is an inappropriate threatened harm for the purposes of reactive defences, as it fails on both grounds. First, chronic pain fails because admission of a defence in such cases would directly frustrate the will of Parliament by undermining the laws prohibiting the possession and supply of controlled/prohibited drugs. In *Attorney General's Reference (No 2 of 2004)*,[56] Lord Mance considered that neither judges nor juries were well-equipped to resolve issues as to when and how far the deliberate policy of legislation should give way to particular instances of hardship, neither knowing what the overall effect of such derogations would be on the legislative scheme as a whole.[57] Indeed, this is a good example of where principle could easily be lost in the pursuit of individual justice. We may feel strong sympathy towards those who struggle with chronic, debilitating pain, but it would be inappropriate to replace one problem with another.[58] The correct route is, as suggested by the courts, reform of the law via Parliament.

However, chronic pain can also be excluded from the category of appropriate threatened harms in reactive defences on logical grounds based on current requirements. In every other example of threatened harm discussed, the harm was extraneous to the accused. We saw in chapter three that currently coercion/necessity require the threatened harm to be extraneous, such that feelings of despair or depression could not ground the defence.[59] In chronic pain situations, while the initial injury may have occurred as a result of an extraneous event, the resulting pain does not, and thus it is more accurate to say that the threat is internal to the accused: it is from their own body that the pain originates. Why is this important? Lord Mance stated that the internal nature of such pain was an issue because the defence an accused relies on must be capable of objective assessment and thus,

---

[55] Although again it is worth remembering that the threatened harm requirement is not the *only* requirement in reactive defence pleas.
[56] [2005] 2 Cr App R 34.
[57] *Ibid* at para [56].
[58] Indeed, claims that denying a necessity defence was akin to subjecting such persons to 'inhuman or degrading treatment' as per Art.3 of the European Convention on Human Rights have also been rejected: *R v Altham* [2006] 2 Cr App R 8.
[59] See section 3.4.3 above.

even on the more lenient interpretation of a 'well-founded belief', the threat must have manifested externally to measure and assess it accordingly.[60] He reached this conclusion by alluding to the case of *Roger and Rose*,[61] where prisoners attempted to escape after finding out their original tariffs had been increased and became suicidal.

While I appreciate the obvious difficulty involved in objectively assessing pain in such situations, I believe this justification is unpersuasive because it risks creating a false equivalence between mental illnesses like depression and chronic pain. It is not clear that neurological issues can easily be equated with the effects of chronic pain, and a ruling which seeks to draw such similarities leaves itself open to potential future appeals, particularly as research in these areas develop. In any case, a more effective logical basis for rejecting chronic pain which avoids this dilemma already exists. The primary reason why the internal aspect of chronic pain is fatal to its inclusion in the threatened harm requirement of reactive defences is that it breaches the characterisation of these defences as a response to an event, as discussed above.[62] To allow a defence in these circumstances would be to stretch beyond recognition our understanding of what it means to react to an event. Indeed, and this relates back to the first point about public policy – if one must continually break the law to achieve an acceptable standard of living, then they do not need a defence, we need a change in the law. For these reasons, and despite the obvious plight of people living with these conditions, it would be inappropriate to include chronic pain in any threatened harm requirement.

9.7 Conclusion

This chapter has sought to explore one of the more controversial aspects of the current reactive defences framework by critically evaluating the threatened harm requirement in the necessity and coercion defences, in terms of its origins and the current criticisms that it faces. I argue that the current demand for threats of death or serious injury is underinclusive, as it does not fully encapsulate the broad concept of personal integrity. I have sought to demonstrate the value of the alternative structure proposed in this thesis at delivering a more refined threatened harm model which can accommodate some of the more nuanced harm situations identified in the courts. The present understanding of the threatened harm requirement has evolved from a symbiotic relationship between the

[60] [2005] 2 Cr App R 34 at para [47].
[61] [1998] 1 Cr App R 143.
[62] See section 9.3.

reactive defences of self-defence and, specifically, coercion in Scots law, resulting in an unprincipled focus on death or serious injury above all else. I have sought to demonstrate that the types of injury that can and should help ground a reactive defence are broader than this, with the requirement being capable of encompassing another infringement to personal autonomy, false imprisonment, without unduly expanding reactive defences beyond reasonable limits. Indeed, it should always be borne in mind that the threatened harm requirement is just one of several, strict requirements limiting reactive defences. I suggest that a normative understanding of coercion, understood as situations of extreme pressure constraining an accused's will, provides an appropriate basis for accepting false imprisonment as a qualifying threatened harm. In contrast, I have drawn the limit at chronic pain, which represents a different kind of challenge – one which is better suited for resolution in Parliament with a change in the law (in terms of our relationship with certain drugs).

10. Conclusion

In this thesis I have sought to critically examine the substantive defences of coercion and necessity in Scots law through various lenses with a view to understanding their nature. As its core premise, this thesis has argued for a normatively focused approach to issues of coercion and necessity in Scots law. Part I of this thesis provided an extensive overview of the defences as they emerged from the early Scottish legal enlightenment in the eighteenth century to contemporary Scots law, with attention paid to the strong influences of English law on their judicial development. This part of the thesis provided the necessary descriptive backdrop with which to frame the rest of the arguments made herein. Of particular importance, I have argued that the law as it has developed in Scotland has created a unified set of requirements for both necessity and coercion, such that the jurisprudence for each defence is mutually applicable. This was largely in part due to the adoption of the English concept of duress of circumstances as our version of necessity in Scotland; a defence which is based on the English variation of coercion, duress by threats. This is a novel way of understanding the way necessity and coercion have developed in Scotland, and perhaps the most controversial claim of this entire thesis.

Having outlined the defences in detail, Part II considered a normative approach to necessity and coercion – one based on underlying values which might be seen as more agreeable than the original, value-laden understanding of necessity rejected by the court in *Dudley and Stephens*. To this end, this thesis has developed two new concepts – situations of individual emergency and situations of extreme pressure – to explain precisely what it is about these reactive defences which exculpates persons who commit an offence when faced with such circumstances. These concepts can be distinguished by their focus on constrained choices and constrained wills respectively, which this thesis argues presents a more coherent dichotomy than is currently recognised between necessity and coercion. Indeed, these terms have been developed, in part, to disassociate from the philosophical baggage attached to the justification/excuse distinction, which often proves to be unhelpful in discussing how these defences should be framed.

Situations of individual emergency, through its factual nexus embodying the idea of a constrained choice, has been defined as covering those situations where a person attempts to bring about a positive result, based on the difficult circumstances they find themselves in. While emphasis is placed on the existence of strenuous circumstances which are

objectively verifiable, our normative assessment focuses on the subjective position of the actor, rather than the normative value of the act itself (such that mistaken beliefs might still ground a defence). The idea of a constrained choice is one which is common in criminal law and, for the most part, the variation presented here is relatively uncontroversial in corresponding to these prior understandings which tend to make up Scots law's understanding of the necessity and coercion defences presently.

In contrast, situations of extreme pressure have been defined as covering those instances where stressful circumstances cause an emotional response in an actor, such that they wrongfully break the law owing to a constrained will. This latter claim required an unpacking of the concept of emotions, both in the law's current understanding of their impact on actions, as well as theories grounded in neuroscience to determine that emotions were an appropriate alternative site for normative blame. It was argued that for a situation of extreme pressure to be made out, the accused's response must be both typical and normatively reasonable, such that it subscribes to the basic values the law seeks to uphold. It is therefore broad enough to encapsulate deserving behaviour which may not otherwise have qualified for a defence based on emotions (i.e. the stoic actor), while being narrow enough to disqualify those emotionally overwhelmed actors whose responses are abhorrent or otherwise unreasonable. In other words, this theory understands emotions to be relevant to exculpation, but they cannot in and of themselves excuse behaviour, particularly the type of behaviour and responses the rest of society would find repugnant.

It is hoped that situations of extreme pressure and situations of individual emergency might form the underlying basis for a new conception of the coercion and necessity defences in Scots law respectively, even if the layman-friendly terminology of 'coercion' and 'necessity' is retained. However, I would recommend that the historical Scots term 'compulsion' should replace 'coercion', given the latter's association with human manipulation which the current thesis has moved away from with respect to the kind of conduct captured by coercion as a defence. On this basis, the necessity defence would consist of an assessment of the accused's choice to determine if it was appropriately constrained and made with the goal of bringing about a positive result. In contrast, compulsion would exculpate based on a constrained will, where a normative assessment of emotion in the circumstances and in relation to the response would determine whether the accused should be held culpable.

In the final part of this thesis, I utilised this normative understanding of these reactive defences to re-examine two of the more controversial requirements of the present

necessity/coercion defences framework – that the threatened harm must be immediate and must be of serious injury or death. These two concepts might be responsible for much of the discussion around coercion and necessity in the current literature, as their strict application seems to fly in the face of logic which suggests that persons might feel genuinely compelled, or act out of necessity, even where no immediate danger exists or they are subject to lesser threats. I have sought to explain, in principled terms, how such requirements should be understood in the new framework.

I have argued that, in the context of temporal requirements where we seek to justify conduct, what matters to reactive defences is the lack of alternative courses of action or, put differently, that the harm threatened was inevitable in the sense that it would have materialised but for the accused's conduct. I have argued that a softer touch might be appropriate where the focus of the defences is on the actions of the accused (in other words, excusatory in nature). For this reason, I suggest that the immediacy requirement should be relaxed such that, while immediate danger would continue to provide a good indicator of credibility, its absence would cease to be fatal to a potential claim of extreme pressure or individual emergency. In practical terms, this would suggest focusing on an assessment of the reasonableness of the accused's response in the circumstances – with an emphasis placed on the reasonableness of an accused's reaction in coercion/compulsion, and on the context of the situation in necessity.

Likewise, in relation to the threatened harm requirement, I have sought to demonstrate that breaches to personal integrity, the very limited reason the law exculpates persons for otherwise breaking the law, is a broader concept than is currently recognised by reactive defences. I have argued that a close historical connection between self-defence and other reactive defences like necessity and coercion has led the latter to broadly adopt similar requirements, without the same underlying rationale. While only the most heinous breaches of personal integrity might ground a plea of (lethal) self-defence, it has been argued that the concept might be modified slightly in coercion/compulsion to include instances of false imprisonment, such that VOT might benefit from a greater protection against prosecution when compelled to commit crime than is currently the case, pursuant to the Warsaw Convention. Nevertheless, on this same normative basis, it was argued that the concept of personal integrity should not extend to chronic pain as the circumstances in such cases engage with another type of situation which goes beyond the more contemporary nature of situations of extreme pressure or individual emergency as envisaged in this thesis.

It is hoped that the content of this thesis can renew the discussion surrounding two of the most elusive substantive defences in Scots law, and perhaps even serve as a foundational point for future arguments and reform, such as answering the question of whether reactive defences other than self-defence ought to apply to charges of homicide. Indeed, in areas like human trafficking where Scots law is still developing appropriate responses, it is vital that our foundational legal concepts can keep pace. In any event, I hope the reader will be convinced that it is high time that Scots law reinjected the foundational sense of normativity currently sorely missing from these defences, grounded in the most basic aspects of the human condition.

Bibliography

Cases

Scottish Cases

Advocate, HM v Anderson (2006) Unreported
Advocate, HM v Greig, Unreported, High Court, May 1979
Advocate, HM v McCallum (1977) SCCR Supp 169
Advocate, HM v Raiker 1989 SCCR 149
Andrew Fairney, *James Purdie and others* (1720), Hume, i, 51
Bone v HM Advocate 2006 SLT 164
Cochrane v HM Advocate 2001 SCCR 655
Collins v HM Advocate 1991 JC 204
Cosgrove v HM Advocate 2008 JC 102
D v Donnelly 2009 SLT 476
Dawson v Dickson 1999 JC 315
Docherty v HM Advocate (1976) SCCR Supp 146
Dolan v McLeod 1999 JC 32
Drury v HM Advocate 2001 SLT 1013
Galbraith v HM Advocate 2002 JC 1
Hai Van Le v HM Advocate [2019] HCJAC 44
Hamilton v Neizer 1993 JC 63
Keay (1837) 1 Swinton 543
Lord Advocate's Reference (No 1 of 2000) 2001 JC 143
Lowe v Mulligan 1991 SCCR 551
Mackay v MacPhail 1989 SCCR 622
McKinnon v HM Advocate 2003 JC 29
McLeod v MacDougall 1989 SLT 151
McNab v Guild 1989 JC 72
Morrison v Valentine 1991 SLT 413
Moss v Howdle 1997 JC 123
Murphy v HM Advocate 2013 JC 60
Owens v HM Advocate 1946 JC 119
Robert Main (1725), Hume, i, 51
Roger Hews (1720), Hume, i, 51
Ruxton v Lang 1998 SCCR 1
Sayers v HM Advocate 1981 SCCR 312
SB v HM Advocate 2015 JC 289
Sinclair v HM Advocate [2017] HCJAC 88
Thomson v HM Advocate 1983 JC 69
Trotter v HM Advocate 2000 SCCR 968
Tudhope v Grubb 1983 SCCR 350
Van Phan v HM Advocate [2018] HCJAC 7
Watson v Hamilton 1988 SLT 316
William and John Riddell (1681), Hume, i, 50

English and Welsh Cases

Attorney General for Jersey v Holley [2005] 2 AC 580
Attorney-General's Reference (No 2 of 1983) [1984] QB 456

Attorney General's Reference (No 2 of 2004) [2005] 2 Cr App R 34
Gillick v West Norfolk and Wisbech AHA [1986] AC 112
DPP v Bell [1992] Crim LR 176
DPP v Jones [1990] RTR 33
DPP v Lynch [1975] AC 653
DPP v Mullally [2006] EWCA Crim 3448 (Admin)
DPP for Northern Ireland v Lynch [1975] AC 653
Evans v Wright [1964] 1 WLUK 183
Pipe v DPP [2012] EWHC 1821 (Admin)
Nicklinson v Ministry of Justice [2014] 2 All ER 32
R (on the application of Nicklinson) v Ministry of Justice [2012] EWHC 2381 (Admin)
R v Abdul-Hussain [1999] Crim LR 570
R v Altham [2006] 2 Cr App R 8
R v Asmelash (Dawit) [2014] QB 103
R v Baines (1970) 54 Cr. App. R. 481
R v Baker and Ward (1999) 2 Cr App R 335
R v Batchelor [2013] EWCA Crim 2638
R v Brandford [2017] 4 WLR 17; [2017] Crim LR 554
R v Clinton [2013] QB 1
R v Cole [1994] Crim LR 582
R v Conway [1989] QB 290
R v Dawes [2014] 1 WLR 947
R v Denton (1987) 85 Cr App R 246
R v Dudley and Stephens (1884) 14 QBD 273
R v Fagossa (1551) 1 Plowd. 1; 75 Eng. Rep. 1
R v Gotts [1992] 2 AC 412
R v Graham (Paul) [1982] 1 WLR 294
R v GS [2018] EWCA Crim 1824
R v Harmer [2001] EWCA Crim 2930
R v Hasan [2005] 2 AC 467
R v Heath [2000] Crim LR 109
R v Howe [1987] AC 417
R v Hudson [1971] 2 QB 202
R v Hussain [2008] EWCA Crim 1117
R v Ireland; R v Burstow [1998] AC 147
R v Jones [2005] Crim LR 122
R v Lewis (1993) 96 Cr App R 412
R v LM, MB, DG, BT and YT [2010] EWCA Crim 2327
R v Lyness [2002] EWCA 1759
R v Martin (1989) 88 Cr App R 343
R v N [2007] EWCA Crim 3479
R v Pommell [1995] 2 Cr App R 607
R v Quayle; R v Wales; R v Taylor and another; R v Kenny [2005] 2 Cr App R 34
R v Rejmanski (Bartosz) [2018] 1 Cr App R 18
R v Rodger and Rose [1998] 1 Cr App R 143
R v S(C) [2012] 1 Cr App R 31
R v Safi [2004] 1 Crim App R 14
R v Sasikaran Selvaratnam [2006] EWCA Crim 1321
R v Sharp (1987) 85 Cr App R 207
R v Shayler [2001] EWCA Crim 1977; [2001] 1 WLR 2206
R v Shortland [1996] 1 Crim App R 116
R v Steane [1947] 1 All ER 813

R v Thacker [2021] 1 Cr App R 21
R v True [2003] EWCA Crim 2255
R v Willer (1986) 83 Cr App R 225
Re A (Children) [2001] Fam 147
Re F (Mental Patient Sterilisation) [1990] 2 AC 1
Santos v CPS Appeals Unit [2013] EWHC 550 (Admin)

Irish Cases

Attorney General v Whelan [1934] IR 518
R v Fegan [1972] NI 80

Canadian Cases

Morgantaler v The Queen [1976] 1 SCR 616
R v Creighton [1993] 3 SCR 3
R v Foster 2018 ONCA 53
R v Hibbert [1995] 2 SCR 973
R v Latimer [2001] SCC 1
R v Lavallee [1990] 1 SCR 852
R v McRae (2005) 199 CCC (3d) 536
R v Mena (1987) 34 CCC (3d) 304
R v Nwanebu 2014 BCCA 387
R v Perka [1984] 2 SCR 232
R v Pétel [1994] 1 SCR 3
R v Ruzic [2001] 1 SCR 687
R v Ryan [2013] SCC 3

US cases

Bram v United States 168 US 532 (1897)
Culombe v Connecticut 367 US 568 (1961)
Leeper v Beltrami 53 Cal. 2d 195, 347 P.2d 12, 1 Cal. Rptr. 12 (1959)
Mincey v Arizona 437 US 385 (1978)
People v Brown, 333 NYS 2d 342 (Sup. Ct. 1972)
State v Norman 324 NC 253, 378 S.E.2d 8 (1989)
State v Hundley, 693 P.2d 475 (Kan. 1985)
State v Johnson 956 F.2d 894, 899 (9th Cir. 1992)
State v Schroeder, 261 N.W.2d 759, 760 (Neb, 1978)
State v Stewart 243 Kan. 639, 763 P.2d 572 (1988)
United States v Ashton 24 F. Cas. 873 (CCD Mass 1834) (No 14,470)
United States v Holmes (1842) 26 F.Cas. 360

Australian Cases

Collingburn (1985) A Crim R 294
Van Den Hoek v R [1986] HCA 76

South African Cases

R v Damascus 1965 (4) SA 598 (SR)
R v Hercules 1954 (3) SA 826 (AD)
R v Mahomed 1938 AD 30
S v Arnold 1985 (3) SA 256 (C)

S v Bailey 1982 (3) SA 772 (A)
S v Chretien 1981 (1) SA 1097 (A)
S v Eadie 2002 (3) SA 719 (SCA)
S v Goliath 1972 (3) SA 1 (A)
S v Laubscher 1988 (1) SA 163
S v Makwayane 1995 6 BCLR 665 (CC)
S v Moses 1996 (1) SACR 701 (C)
S v Nursingh 1995 (2) SACR 331 (D)
S v Oosthuizen 2018 JDR 0725 (SCA)
S v Potgieter 1994 (1) SACR 61 (A)
S v van Vuuren 1983 (1) SA 12 (A)
S v Wiid 1990 (1) SACR 561 (A)

Legislation and Treaties

UK Legislation

Coroners and Justice Act 2009, *s54(1), s54(1)(a), s54(1)(b), s54(1)(c), s55, s55(3), s55(4), s55(5), s55(6)(c), s56*

Criminal Justice and Immigration Act 2008, *s76*

Homicide Act 1957, *s3*

Human Rights Act 1998, *s1(1)(a), s1(2), s6(1), s6(3)*

Human Trafficking and Exploitation (Scotland) Act 2015, *s8*

Modern Slavery Act 2015, *s45(1)*

Road Traffic Offenders Act 1988, *s34*

European Legislation and Conventions

Convention for the Protection of Human Rights and Fundamental Freedoms, *Article 3*

Directive 2011/36/EU of the European Parliament and of the Council of 5 April 2011 on preventing and combating trafficking in human beings and protecting its victims, and replacing Council Framework Decision 2002/629/JHA, *recital 14, Article 26*

Canadian Legislation

Canadian Criminal Code (RSC 1985, C-46), *s17, s34, s34(2)(b)*

Constitution Act (1982) (Canadian Charter of Rights and Freedoms), *s7*

The Citizen's Arrest and Self-Defence Act 2012, SC 2012, C-9

South African Legislation

Criminal Procedure Act 51 of 1977, *s78(7)*

Articles & Book Chapters

Akande, D & Liefländer, T, "Clarifying Necessity, Imminence, and Proportionality in the Law of Self-Defense" (2013) 107(3) *American Journal of International Law* 563

Alexander, L, "Lesser Evils: A Closer Look at the Paradigm Justification" (2005) 24 *Law and Philosophy* 611

Allhoff, F, "Self-defense Without Imminence" (2019) 56 *American Criminal Law Review* 1527

Ammons, LL, "Discretionary Justice: A Legal and Policy Analysis of a Governor's Use of the Clemency Power in the Case of Incarcerated Battered Women" (1994) 3(1) *Journal of Law and Policy* 1

Appel, SD, "Beyond Self-Defense: The Use of Battered Woman Syndrome in Duress Defences" (1994) 4 *University of Illinois Law Review* 955

Armour, JD, "*Race Ipsa Loquitur:* Of Reasonable Racists, Intelligent Bayesians, and Involuntary Negrophobes" (1994) 46 *Stanford Law Review* 781

Arnolds, EB & Garland, NF, "The Defence of Necessity in Criminal Law: The Right to Choose the Lesser Evil" (1974) 65 *Journal of Criminal Law & Criminology* 289

Baron, M, "Self-Defense: The Imminence Requirement" in L Green & B Leiter (eds), *Oxford Studies in Philosophy of Law: Volume 1* (New York: Oxford University Press, 2011) pp.228-266

Bauer, SM & Eckerstrom, PJ, "The State Made Me Do It: the Applicability of the Necessity Defence to Civil Disobedience" (1987) 39 *Stanford Law Review* 1173

Bayles, MD, "Character, Purpose, and Criminal Responsibility" (1982) 1 *Law and Philosophy* 5

Bayles, MD, "Reconceptualising Necessity and Duress" (1987) 33 *Wayne Law Review* 1191

Berger, BL, "Emotions and the Veil of Voluntarism: The Loss of Judgment in Canadian Criminal Defences" (2006) 51 *McGill Law Journal* 99

Bond, J & Fougere, M, "Omnipresent Threats: A Comment on the Defence of Duress in International Criminal Law" (2014) 14 *International Criminal Law Review* 471

Botterell, A, "Understanding the Voluntary Act Principle" in F Tanguay-Renaud & J Stribopoulos (eds), *Rethinking Criminal Law Theory: New Canadian Perspectives* (Oxford: Hart Publishing, 2012), pp.97-115

Brudner, A, "A Theory of Necessity" (1987) 7 *Oxford Journal of Legal Studies* 339

Burchell, J, "A Provocative Response to Subjectivity in the Criminal Law" 2003 *Acta Juridica* 23

Burke, AS, "Rational Actors, Self-Defense, and Duress: Making Sense, Not Syndromes, out of the Battered Woman" (2002) 81(1) *North Carolina Law Review* 211

Carr, CL, "Duress and Criminal Responsibility" (1991) 10(2) *Law and Philosophy* 161

Chan, W & Simester, AP, "Duress, Necessity: How Many Defences? (2005) 16 *King's Law Journal* 121

Chiao, V, "Action and Agency in the Criminal Law" (2009) 15 *Legal Theory* 1

Clarkson, CMV, "Necessary Action: A New Defence" [2004] Crim LR 81

Coughlan, S, Ferguson, G, Stuart, D, Berger, B, Mathen, C & Sankoff, P, "Reform of the Defence of Duress (and Necessity)" (2018) 66 *Criminal Law Quarterly* 229

Dowd, M, "Dispelling the Myths About the 'Battered Woman's Defense:' Towards a New Understanding" (1992) 19 *Fordham Urban Law Journal* 567

Dressler, J, "Battered Women and Sleeping Abusers: Some Reflections" (2006) 3 *Ohio State Journal of Criminal Law* 457

Dressler, J, "Battered Women Who Kill Their Sleeping Tormentors" in A Simester & S Shute (eds), *Criminal Law Theory: Doctrines of the General Part* (London: Oxford University Press, 2002) pp.259-282

Dressler, J, "Duress" in J Deigh & D Dolinko, *The Oxford Handbook of Philosophy of Criminal Law* (New York: Oxford University Press, 2011), pp.269-298

Dressler, J, "Exegesis of the law of Duress: Justifying the Excuse and Searching for its Proper Limits" (1989) 62 *Southern California Law Review* 1331

Dsouza, M, "Retreat, Submission, and the Private Use of Force" (2015) 35 *Oxford Journal of Legal Studies* 727

Duff, RA, "Choice, Character, and Criminal Liability" (1993) 12 *Law and Philosophy* 345

Farmer, L, "The Idea of Principle in Scots Criminal Law" in J Chalmers, F Leverick & L Farmer (eds), *Essays in Criminal Law in Honour of Sir Gerald Gordon* (Edinburgh University Press, 2010), pp.86-102

Ferguson, PW, "Necessity and Duress in Scots Law" [1986] Crim LR 103

Ferzan, KK, "Defending Imminence: From Battered Women to Iraq" (2004) 46(2) *Arizona Law Review* 213

Fingarette, H, "Victimization: A Legalist Analysis of Coercion, Deception, Undue Influence, and Excusable Prison Escape" (1985) 42 *Washington and Lee Law Review* 65

Finkelstein, CO, "Duress: A Philosophical Account of the Defense in Law" (1995) 37 *Arizona Law Review* 251

Finkelstein, CO, "Self-Defense as a Rational Excuse" (1996) 57 *University of Pittsburgh Law Review* 621

Fischer, JM, "Responsiveness and Moral Responsibility" in F Schoeman (ed), *Responsibility, Character, and the Emotions* (Cambridge: Cambridge University Press, 1987) pp.81-106

Fletcher, GP, "Domination in the Theory of Justification and Excuse" (1996) 57 *University of Pittsburgh Law Review* 553

Fletcher, GP, "The Individualization of Excusing Conditions" (1974) 47 *Southern California Law Review* 1269

Foot, P, "The Problem of Abortion and the Doctrine of Double Effect" in *Virtues and Vices and Other Essays in Moral Philosophy* (Oxford: Basil Blackwell, 1978) (reprinted from (1967) 5 *The Oxford Review* 5) pp.19-31

Gardner, J, "The Gist of Excuses" (1998) 1 *Buffalo Criminal Law Review* 575

Gauthier, D, "Self-Defense and the Requirement of Imminence: Comments on George Fletcher's *Domination in the Theory of Justification and Excuse*" (1996) 57 *University of Pittsburgh Law Review* 615

Glazebrook, PR, "The Necessity Plea in English Criminal Law" (1972) 30(1) *Cambridge Law Journal* 87

Goosen, S, "Battered Women and the Requirement of Imminence in Self-Defence" (2013) 16(1) *Potchefstroom Electronic Law Journal* 70

Greenawalt, K, "The Perplexing Borders of Justification and Excuse" (1984) 84 *Columbia Law Review* 1897

Guz, A & McMahon, M, "Is Imminence Still Necessary? Current Approaches to Imminence in the Laws Governing Self-Defence in Australia" (2011) 13(2) *Flinders Law Journal* 79

Hoctor, S, "Non-Pathological Criminal Incapacity Relating to Provocation or Emotional Stress – An Overview of Developments in South African Law" (2019) 49 *South African Journal of Psychology* 177

Horder, J, "Criminal Culpability: The Possibility of a General Theory" (1993) 12 *Law and Philosophy* 193

Horder, J, "Self-Defence, Necessity and Duress: Understanding the Relationship (1998) 11 *Canadian Journal of Law and Jurisprudence* 143

Huigens, K, "Virtue and Inculpation" (1995) 108 *Harvard Law Review* 1423

Ireland, E, "Rebutting the Presumption: Rethinking the Common Law Principle of Marital Coercion in Eighteenth- and Nineteenth-Century England" (2019) 40 *Journal of Legal History* 21

James, W, "What is an Emotion?" in CG Lange & W James, *The Emotions*, *Vol. 1* (Baltimore; Williams & Wilkins Co, 1922) pp.11-30

Kadish, SH, "Excusing Crime" (1987) 75 *California Law Review* 257

Kahan, DM, "The Secret Ambition of Deterrence" (1999) 113 *Harvard Law Review* 413

Kahan, DM & Nussbaum, MC, "Two Conceptions of Emotion in Criminal Law" (1996) 96 *Columbia Law Review* 269

Karstedt, S, "Emotions and Criminal Justice" (2002) 6(3) *Theoretical Criminology* 299

Kaufman, WRP, "Self-Defense, Imminence, and the Battered Woman" (2007) 10(3) *New Criminal Law Review: An International and Interdisciplinary Journal* 342

Kennedy, C, "Defences: Justification, Excuse and Provocation" in M Hill, N Doe, RH Helmholz & J Witte (eds), *Christianity and Criminal Law: An Introduction* (Oxford: Routledge, 2020), pp.253-68

Klimchuk, D, "Aristotle on Necessity and Voluntariness" (2002) 19 *History of Philosophy Quarterly* 1

Klimchuk, D, "Moral Innocence, Normative Involuntariness, and Fundamental Justice" (1998) 18 *Criminal Reports* (5th) 96

Laird, K, "Evaluating the Relationship between Section 45 of the Modern Slavery Act 2015 and the Defence of Duress: An Opportunity Missed?" [2016] Crim LR 395

Lord Kilbrandon, "Duress *Per Minas* as a Defence to Crime: I" (1982) 1(2) *Law and Philosophy* 185

Louw, R, "*S v Eadie*: The End of the Road for the Defence of Provocation?" (2003) 16 *South African Journal of Criminal Justice* 200

Mackay, RD & Mitchell, BJ, "Provoking Diminished Responsibility: Two Pleas Merging into One?" [2003] Crim LR 745

MacLean, J, Verrelli, N & Chambers, L, "Battered Women Under Duress: The Supreme Court of Canada's Abandonment of Context and Purpose in *R v Ryan*" (2017) 29(1) *Canadian Journal of Women and the Law* 60

Maguigan, H, "Battered Women and Self-Defense: Myths and Misconceptions in Current Reform Proposals (1991) 140(2) *University of Pennsylvania Law Review* 379

Mahoney, M, "Legal Images of Battered Women: Redefining the Issue of Separation (1991) 90(1) *Michigan Law Review* 1

McColgan, A, "In Defence of Battered Women Who Kill (1993) 13(4) *Oxford Journal of Legal Studies* 508

McPherson, R, "Legal change and legal inertia: understanding and contextualising Scottish cases in which women kill their abusers" (2021) 5 *Journal of Gender-Based Violence* 289

Moriarty, JC, "While Dangers Gather: The Bush Preemption Doctrine, Battered Women, Imminence, and Anticipatory Self-Defense" (2005) 30(1) *New York University Review of Law & Social Change* 1

Murdoch, JB, "Is Imminence Really Necessity? Reconciling Traditional Self-Defense Doctrine With The Battered Woman Syndrome" (2000) 20(1) *Northern Illinois University Law Review* 191

Nourse, V, "Passion's Progress: Modern Law Reform and the Provocation Defense" (1997) 106(5) *Yale Law Journal* 1331

Nourse, V, "The New Normativity: The Abuse Excuse and the Resurgence of Judgment in the Criminal Law" (1998) 50(4) *Stanford Law Review* 1435

Ormerod, D, "Duress: *R v van Dao (Vinh)*" (case commentary) [2013] Crim LR 234

Paley, JM, "Compulsion: Fear and the Doctrine of Necessity" 1971 *Acta Juridica* 205

Parry, JT, "The Virtue of Necessity: Reshaping Culpability and the Rule of Law" (1999) 36 *Houston Law Review* 397

Reed, A, "Duress and Specificity of Threat: Commentary on *R v Z*" (2003) 67 *Journal of Criminal Law* 281

Reilly, A, "The Heart of the Matter: Emotion in Criminal Defences" (1997) 29 *Ottawa Law Review* 117

Reilly, A & Mikus, R, "R v Hibbert: The Theoretical Foundations of Duress" (1996) 30 *University of British Columbia Law Review* 181

Robinson, OF, "Law, Morality and Sir George Mackenzie" in HL MacQueen (ed), *Miscellany Six* (Stair Society, 2009), pp.11-27

Robinson, PH, "Competing Theories of Justification: Deeds v. Reasons" in AP Simester & ATH Smith (eds), *Harm and Culpability* (Oxford: Oxford University Press, 1996) pp.45-70

Robinson, PH, "Criminal Law Defenses: A Systematic Analysis" (1982) 82 *Columbia Law Review* 199

Robinson, S, "Necessity's Newest Inventions" (1991) 11(1) *Oxford Journal of Legal Studies* 125

Rosen, CJ, "The Excuse of Self-Defense: Correcting A Historical Accident on Behalf of Battered Women Who Kill" (1986) 36 *American University Law Review* 11

Rosen, R, "On Self-Defense, Imminence, and Women Who Kill Their Batterers" (1993) 71(2) *North Carolina Law Review* 371

Rubin, G, "Pre-dating Vicky Pryce: The Peel Case (1922) and the Origins of the Marital Coercion Statutory Defence" (2014) 34 *Legal Studies* 631

Sabini, J & Silver, M, "Emotions, Responsibility and Character" in F Schoeman (ed), *Responsibility, Character, and the Emotions* (Cambridge University Press, 1987) pp.164-75

Schulhofer, SJ, "Harm and Punishment: A Critique of Emphasis on the Results of Conduct in the Criminal Law" (1974) 122 *University of Pennsylvania Law Review* 1497

Simpson, B, "Modern Slavery and Prosecutorial Discretion: When is it in the Public Interest to Prosecute Victims of Trafficking?" (2019) 83(1) *J. Crim. L.* 14

Snyman, CR, "Criminal Justice in a New Society" 2003 *Acta Juridica* 1

Snyman, CR, "Is There Such a Defence as Emotional Stress?" (1985) 102 *South African Law Journal* 240

Stark, F, "Necessity and *Nicklinson*" [2013] *Criminal Law Review* 949

Tadros, V, "The Characters of Excuse" (2001) 21 *Oxford Journal of Legal Studies* 495

Tadros, V, "The Structure of Defences in Scots Criminal Law" (2003) 7 *Edinburgh Law Review* 60

Tanguay-Renaud, F, "Individual Emergencies and the Rule of Criminal Law" in F Tanguay-Renaud & J Stribopoulos (eds), *Rethinking Criminal Law Theory: New Canadian Perspectives* (Oxford: Hart Publishing, 2012), pp.21-54

Taurek, JM, "Should the Numbers Count?" (1977) 6(4) *Philosophy and Public Affairs* 293

Thomson, JJ, "Killing, Letting Die, and the Trolley Problem" (1976) 59(2) *The Monist* 204

Thomson, JJ, "The Trolley Problem" (1985) 94(6) *The Yale Law Journal* 1395

Thomson, JJ, "Turning the Trolley" (2008) 36(4) *Philosophy and Public Affairs* 359

Thorburn, M, "Justifications, Powers and Authority" (2008) 117 *Yale Law Journal* 1070

Trotter, GT, "Necessity and Death: Lessons from *Latimer* and the Case of the Conjoined Twins" (2003) 40 *Alberta Law Review* 817

Wallace, S, "Beyond Imminence: Evolving International Law and Battered Women's Right to Self-Defense" (2004) 71 *The University of Chicago Law Review* 1749

Warner, CT, "Anger and Similar Delusions" in R Harré (ed), *The Social Construction of Emotions* (Oxford: Blackwell, 1986), pp.135-66

Wiederman, M, "Premarital Sex" in *Sex and Society,* Vol.3 (New York: Marshall Cavendish Reference, 2010) pp.663-66

Williams, G, "Necessity: Duress of Circumstances or Moral Involuntariness?" (2014) 43 *Common Law World Review* 1

Willoughby, MJ, "Rendering Each Woman Her Due: Can a Battered Woman Claim Self-Defense When She Kills Her Sleeping Batterer" (1989) 38(1) *University of Kansas Law Review* 169

Yaffe, G, "Prevention and Imminence, Pre-Punishment and Actuality" (2011) 48(4) *San Diego Law Review* 1205

Yeo, S, "Compulsion and Necessity in African Criminal Law" (2009) 53 *Journal of African Law* 90

Books & Theses

Alexander, L & Ferzan, KK, *Reflections on Crime and Culpability: Problems and Puzzles* (Sheridan Books, Inc. & Cambridge University Press, 2018)

Alison, A, Sir, *Principles of the Criminal Law of Scotland* (Edinburgh: William Blackwood, 1832)

Anderson, AM, *The Criminal Law of Scotland* (Edinburgh: Bell & Bradfute, 1892)

Annas, J, *Intelligent Virtue* (Oxford: Oxford University Press, 2011)

Aristotle, *The Nicomachean Ethics*, translation by D Ross (Oxford: Oxford University Press, 2009)

Bacon, F, Sir, *The Elements of the Common Lawes of England*, Volume 13 (London: Assignes of I. More Esq, 1630)

Bakircioglu, O, *Self-Defence in International and Criminal Law: The Doctrine of Imminence* (Oxford: Routledge, 2011)

Balfour, J, Sir, *The Practicks of Sir James Balfour of Pittendreich* (1754) (PGB McNeill (ed) 1962, Vol.21-22, The Stair Society)

Barclay, G, "The Structure of Assault in Scots Law: A Historical and Comparative Perspective" (2017) University of Glasgow LLM(R) Thesis: http://theses.gla.ac.uk/8569/

Blackstone, W, Sir, *Commentaries on the Laws of England*, Vol. IV (Oxford: Clarendon Press, 1769)

Burchell, J, *Principles of Criminal Law* (3rd revised ed, South Africa: Juta & Co, 2006)

Chalmers, J & Leverick, F, *Criminal Defences and Pleas in Bar of Trial* (Edinburgh: W Green & Son Ltd, 2006)

Conrad, CS, *Jury Nullification: The Evolution of a Doctrine* (Washington, DC: Cato Institute Press, 1998)

Dershowitz, A, *The Abuse Excuse: And Other Cop-Outs, Sob Stories, and Evasions of Responsibility* (Boston, MA: Little Brown & Co., 1994)

Dsouza, M, *Rationale-Based Defences in Criminal Law* (Hart Publishing, 2019)

Dubber, MD, *An Introduction to the Model Penal Code* (New York: Oxford University Press, 2015)

Duff, RA, *Answering for Crime* (Oxford: Hart Publishing, 2009)

Farmer, L, *Making the Modern Criminal Law: Criminalization and Civil Order* (Oxford: Oxford University Press, 2016)

Ferguson, PR & McDiarmid, C, *Scots Criminal Law: A Critical Analysis* (2nd edn, Edinburgh: Edinburgh University Press, 2014)

Fletcher, GP, *Rethinking Criminal Law* (New York: Oxford University Press, 2000)

Glare, PGW, *Oxford Latin Dictionary* (London: Clarendon Press, 1968)

Gordon, GH, Sir, *Criminal Responsibility in Scots Law* (1960) PhD Thesis: https://theses.gla.ac.uk/2753/1/1960gordonphd.pdf

Gordon, GH, Sir, *The Criminal Law of Scotland,* Vol I (3rd ed, MGA Christie (ed), Edinburgh: W Green & Son Ltd, 2000)

Green, TA, *Verdict According to Conscience: Perspectives on the English Criminal Trial Jury 1200-1800* (Chicago & London: University of Chicago Press, 1985)

Hale, M, Sir, *The History of the Pleas of the Crown*, Volume 1, (London: Printed by E & R Nutt, and R Gosling (the assigns of Edward Sayer), for F Gyles, T Woodward, and C Davis, 1736)

Harré, R, *Social Being* (New Jersey, Littlefield: Adams & Co, 1980)

Hart, HLA, *Punishment and Responsibility*, (2nd edn, Oxford: Oxford University Press, 2008)

Hirstein, W, Sifferd, KL & Fagan, TK, *Responsible Brains: Neuroscience, Law, and Human Culpability* (Cambridge, MA: MIT Press, 2018)

Hobbes, T, *Leviathan* (London: Andrew Crooke, 1651)

Horder, J, *Provocation and Responsibility* (Oxford: Clarendon Press, 1992)

Kadish, MR & Kadish, SH, *Discretion to Disobey: A Study of Lawful Departures from Legal Rules* (Stanford, California: Stanford University Press, 1973)

Lacey, N, *State Punishment* (London: Routledge, 1988)

Lee, C, *Murder and the Reasonable Man: Passion and Fear in the Criminal Courtroom* (New York: New York University Press, 2003)

Leverick, F, *Killing in Self-Defence* (Oxford: Oxford University Press, 2006)

Lorenz, K, *On Aggression*, translation by MK Wilson (New York, Harcourt: Brace & World Inc, 1963)

Macdonald, JHA, *A Practical Treatise on the Criminal Law of Scotland* (Edinburgh: William Paterson, 1867)

Macdonald, JHA, *A Practical Treatise on the Criminal Law of Scotland* (2nd edn, Edinburgh: William Paterson, 1877)

Macdonald, JHA & Macdonald, ND, *A Practical Treatise on the Criminal Law of Scotland* (3rd edn, Edinburgh: W Green & Son Ltd, 1894)

Macdonald, JHA, *A Practical Treatise on the Criminal Law of Scotland* (RM Mitchell (ed), 4th edn, Edinburgh: W Green & Son Ltd, 1929)

Macdonald, JHA, *A Practical Treatise on the Criminal Law of Scotland* (J Walker & DJ Stevenson (eds), 5th edn, Edinburgh: W Green & Son Ltd, 1948)

Mackenzie, G, Sir *The Laws and Customs of Scotland in Matters Criminal* (1678) (OF Robinson (ed), Vol.59, The Stair Society, 2012)

McMahan, J, *Killing in War* (New York: Oxford University Press, 2009)

Menochius, J, *De arbitrariis judicum quaestionibus & causis libri duo quibus tota fere iuris pars quae a iudicum arbitrio & potestate pendet ... pertractatur ... : accessit postremae huic editioni eiusdem argumenti centuria sexta antea quidem a D. auctore conscripta sed nunc denuo emendatius excusa* (Ex Officina Antonij Hierati, 1607)

Menochius, J, *De arbitrariis judicum quaestionibus & causis libri duo quibus tota fere iuris pars quae a iudicum arbitrio & potestate pendet ... pertractatur ... : accessit postremae huic editioni eiusdem argumenti centuria sexta antea quidem a D. auctore conscripta sed nunc denuo emendatius excusa* (Coloniae Agrippinae, 1613)

Moore, MS, *Placing Blame: A Theory of the Criminal Law* (New York: Oxford University Press, 2010)

Norrie, A, *Crime, Reason and History: A Critical Introduction to Criminal law*, (3rd edn, London: Cambridge University Press, 2014)

Nozick, R, *Philosophical Explanations* (Oxford: Clarendon Press, 1981)

Nussbaum, MC, *The Therapy of Desire: Theory and Practice in Hellenistic Ethics* (Princeton, NJ: Princeton University Press, 1994)

Rawls, J, *Political Liberalism* (New York: Columbia University Press, 1996)

Robinson, PH, *Criminal Law Defenses* (July 2021 Update, Thomson Reuters) (Minnesota: West Publishing Co, 1984)

Rodin, D, *War and Self-Defense* (Oxford: Clarendon Press, 2002)

Simpson, AWB, *Cannibalism and the Common Law: The Story of the Tragic Last Voyage of the Mignonette and the Strange Legal Proceedings to Which it Gave Rise* (Harmondsworth: Penguin, 1986)

Skene, J, Sir *Regiam Majestatem: The Auld Lawes and Constitutions of Scotland* (Edinburgh: Thomas Finlason, 1609)

Smith, A, *The Theory of Moral Sentiments,* (1759) (DD Raphael & AL Macfie (eds), Indianapolis: Liberty Classics, 1982)

Solomon, R, *The Passions* (Notre Dame: University of Notre Dame Press, 1976)

Spain, E, *The Role of Emotions in Criminal Law Defences: Duress, Necessity and Lesser Evils* (Cambridge: Cambridge University Press, 2011)

Uniacke, S, *Permissible Killing: The Self-Defence Justification of Homicide* (Cambridge: Cambridge University Press, 1994)

Weber, M & Winckelmann, J, *Gesammelte Politische Schriften*, (5th edn, Tübingen: Mohr, 1988)

Wertheimer, A, *Coercion* (New Jersey: Princeton University Press, 1987)

Williams, GL, *Criminal Law: The General Part* (1st edn, London: Stevens & Sons; 1953)

Williams, GL, *Textbook of Criminal Law* (2nd edn, London: Stevens, 1983)

Wilson, JQ, *Moral Judgment: Does the Abuse Excuse Threaten our Legal System?* (New York: Harper Collins, 1997)

Yaffe, G, *Attempts* (New York: Oxford University Press, 2010)

Yeo, S, *Compulsion in the Criminal Law* (New South Wales: Law Book Company, 1990)

Reports and other Resources

American Law Institute, Model Penal Code (citing §3.04(1), §210.3)

Crown Office & Procurator Fiscal Service, *Lord Advocate's Instructions for Prosecutors when considering Prosecution of Victims of Human Trafficking and Exploitation* (August 2021) (accessible at: https://www.copfs.gov.uk/publications/prosecution-policy-and-guidance?showall=&start=4)

European Research Council - *The Construction of Objectivity: An international perspective on the emotive-cognitive process of judicial decision-making (JUSTEMOTIONS)*

Law Commission, *Murder, Manslaughter and Infanticide* (LC 304) (2006)

Law Reform Commission of Ireland, Consultation Paper: "Duress and Necessity" (Ireland: LRC CP 39) (2006)

New Zealand Law Commission, *Some Criminal Defences with Particular Reference to Battered Defendants* (R73, 22 May) (Wellington: NZLC, 2001)

Scottish Law Commission, *Seventh Programme of Law Reform* (Scot Law Com No 198) (2005)