

**A CRITICAL ANALYSIS OF THE CURRENT APPROACH OF THE
COURTS AND ACADEMICS TO THE PROBLEM OF EVIDENTIAL
UNCERTAINTY IN CAUSATION IN TORT LAW**

By
Gemma Turton

A thesis submitted to the
University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

School of Law
College of Arts, Humanities, and Law
University of Birmingham
October 2012

**UNIVERSITY OF
BIRMINGHAM**

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

The primary aim of this thesis is to identify a coherent legal response to the particular causal problem of the ‘evidentiary gap’. In order to do this, it is necessary to understand how the ‘evidentiary gap’ relates to causation in negligence more generally, so the thesis addresses both the nature and function of the tort of negligence as well as the role played by causation within that tort. It argues that negligence is best understood as a system of corrective justice-based interpersonal responsibility. In this account, causation has a vital role so the test of causation must be philosophically sound. Causation, however, also occupies only a limited role so analysis must draw fully on the doctrines of damage and breach which bracket the causation inquiry, as well as notions of quantification of loss. The NESS test for causation is shown to be preferable to the but-for test because it is conceptually more adequate and therefore able to address causal problems that the but-for test cannot. This thesis rejects claims for proportionate recovery based on the notion of loss of a chance of avoiding physical harm in medical negligence, but proposes limited recovery for loss of a chance as an independent form of damage arising because of unique considerations of interpersonal responsibility in the doctor-patient relationship in cases of misdiagnosis/mistreatment of existing illness. It is argued that the *Fairchild* test of material contribution to risk of harm in cases of evidentiary gap is not consistent with corrective justice, and that this cannot be resolved by reconceptualising the gist of the action as the risk of harm. The *Fairchild* exception lacks coherence because of its instrumentalist basis, so should not be applied outside of the mesothelioma context.

Acknowledgments

I would like to thank my supervisor, Claire McIvor, for all her help and support throughout the time it has taken to research and write this thesis. She has not only guided me academically but has also developed my confidence and self-belief. I owe her an enormous debt of gratitude. I would also like to thank Sarah Green for inspiring my interest in causation as an undergraduate student. I am grateful to the University of Birmingham for giving me the postgraduate teaching assistantship that made this financially possible, and to the University of Leicester for giving me a period of study leave. I have been fortunate to present papers on aspects of this research, so I would like to thank the School of Law and the Institute of Medical Law at the University of Birmingham, the School of Law at the University of Leicester, the European Association of Health Law, and the Obligations VI conference for allowing me this opportunity. A number of people have read work on aspects of the thesis, and I would particularly like to thank James Lee, John Hartshorne, and Jose Miola for this.

I would like to thank a number of my friends who have been incredibly supportive (and patient), especially Martin George, Laura Brampton, Sophie Boyron, Sally Cunningham, Daniel Attenborough, and Paul O'Connell. I owe great thanks to my parents for their encouragement. Most of all, I would like to thank my sister, Samantha, for her love and encouragement throughout the years. She is an inspiration in herself and none of this would have been possible without her.

Table of Contents

Introduction.....	1
Chapter One: Theoretical and doctrinal framework.....	6
1. The theoretical basis of negligence.....	7
1.1 The Aristotelian conception of corrective justice.....	11
1.1.1 Justice as a whole.....	11
1.1.2 Justice as a part: corrective and distributive justice.....	12
1.1.2.1 Distributive justice.....	13
1.1.2.2 Corrective justice.....	14
1.2 The moral basis of corrective justice: Aristotelian corrective justice and Kantian Right	17
1.2.1 Kantian Right.....	19
1.3 Incoherence of consequentialist theories	22
1.4 Moving beyond formalism to the legal system - combining Kantian Right and societal morality.....	26
1.5 The place of causation in corrective justice-based interpersonal responsibility	32
1.5.1 The pivotal place of causation in corrective justice.....	32
1.5.2 The limited role of causation.....	35
2. The doctrinal framework.....	38
2.1 The duty of care	39
2.1.1 Duty as the ‘central organising concept’.....	39
2.1.2 Identifying issues of interpersonal responsibility in the duty doctrine.....	42
2.2 Actionable damage/the gist of the negligence action.....	48
2.2.1 Damage.....	49
2.2.2 Quantification	49
2.3 Breach of duty	52
Chapter Two: Identifying the Proper Function of Causation.....	55
1. The demands on causation.....	57

1.1	The limited role of causation in negligence	57
1.2	The robustness of the philosophical account.....	61
2.	The philosophical account of causation.....	63
2.1	Hume: contiguity, priority, constant conjunction and necessity.....	63
2.2	Mill: the addition of ‘plurality’ and ‘complexity’.....	65
	2.2.1 Complexity of causes	66
	2.2.2 Plurality of causes.....	68
3.	Tests for causation.....	70
3.1	Complexity and plurality of causes as necessary and sufficient conditions.....	70
3.2	But-for and NESS in simple cases	72
3.3	Over-determined causation	74
	3.3.1 But-for, NESS and pre-emption.....	75
	3.3.2 But-for, NESS and duplication	76
4.	Overcoming potential problems with NESS	79
4.1	Defining key causal terms.....	80
4.2	The problem of asymmetrical duplication	82
	4.2.1 The solution to asymmetrical duplication.....	83
	4.2.2 Causation of divisible damage	88
	4.2.2.1 Divisible damage and interaction between causes	91
5.	Using NESS to overcome common problems with exceptional legal tests	92
5.1	<i>Bonnington Castings v Wardlaw</i>	94
	5.1.1 The scope of the test	96
	5.1.1.1 Divisibility of damage.....	96
	5.1.1.2 ‘Cumulative’ causal process	98
	5.1.2 Analysis of the <i>Wardlaw</i> test.....	99
5.2	Practical application of the NESS test.....	101
	5.2.1 The question of causation in <i>Bailey</i>	101
	5.2.2 The question of ‘apportionment’	103

Chapter Three: Loss of a chance.....	107
1. Analysis of <i>Hotson v East Berkshire Health Authority</i>	108
1.1 The facts.....	109
1.2 Distinguishing damage, causation, and quantification	111
1.2.1 The ‘hook’ argument.....	112
1.2.2 The vicissitudes principle	114
1.3 The contract/tort distinction	116
2. Analysis of <i>Gregg v Scott</i>	118
2.1 Pain and suffering.....	119
2.2 Loss of life expectancy.....	121
3. Statistics, chance, and the balance of probabilities standard of proof.....	123
3.1 What must the claimant prove? What is the problem of proof?	125
3.1.1 Damage.....	126
3.1.2 Causation	126
3.2 The balance of probabilities standard of proof.....	129
3.3 Proof in <i>Hotson</i>	134
3.4 Proof in <i>Gregg</i>	136
3.4.1 Reliability of the evidence	138
3.4.2 Personalising the evidence	139
4. ‘Loss of a chance’ as damage.....	141
4.1 Rejecting ‘chance’ conceived as a proportion of the physical outcome	142
4.2 Epistemological chance as an independent form of damage	145
4.2.1 Post-diagnosis factors that determine the outcome.....	146
4.2.2 The grey-areas on the deterministic-indeterministic spectrum.....	148
4.2.3 When does a patient have a ‘chance’? Chance vs. risk.....	152
4.2.3.1When does a patient have a meaningful chance?	154
4.2.4 ‘Loss’ of a chance	158
4.2.5 Valuing the chance	160

Chapter Four: The evidentiary gap	164
1. Defining the evidentiary gap	166
1.1 The evidentiary gap relating to dermatitis: <i>McGhee v National Coal Board</i>	167
1.2 The evidentiary gap relating to mesothelioma: <i>Fairchild v Glenhaven Funeral Services</i>	168
1.3 Analysis of the evidentiary gap.....	171
1.3.1 NESS and identifying the problem: locating the evidentiary gap	172
1.3.2 NESS: separating the conceptual from the evidential aspects of the legal solution	174
2. Evidential solutions: the inferential interpretation of <i>McGhee</i>	177
2.1. The inferential approach to <i>McGhee</i>	178
2.1.1 Scientific and legal standards of proof.....	180
2.1.1.1 The place of intuition in the balance of probabilities.....	183
2.1.1.2 Evidence of the evidentiary gap	184
2.2 The ‘doubling of the risk’ approach to proof on the balance of probabilities	187
2.2.1 Doubling of the risk.....	188
2.2.2 Closing the evidentiary gap	191
3. Risk of harm as the gist of the negligence action: <i>Barker v Corus (UK) plc</i>	194
3.1 The decision in <i>Barker v Corus</i>	195
3.2 Analysis of the ‘risk as gist’ approach.....	199
3.2.1 The inconsistency of the physical harm requirement.	201
3.2.2 Physical harm and risk: the shift in perspectives	202
3.2.2.1 What is ‘risk?’.....	203
3.2.2.2 Can risk constitute damage?	205
3.2.2.2.1 Calculating risk and probability.....	206
3.2.2.2.2 ‘Increase in risk’ or ‘contribution to risk?’	208
3.2.3 Confusion in the ‘apportionment’ issue.....	210
3.3 Can risk constitute damage in corrective justice?	211

3.3.1 Objective risk: the evidentiary gap prevents proof that the defendant created an objective risk.....	213
3.3.2 Subjective/epistemic risk: lacks the moral significance to be damage	214
3.3.2.1 The difficulty of explaining why epistemic risk deserves compensation.....	215
3.3.2.2 Epistemic risk: the conflation of breach and damage	218
4. Theoretical approaches to the ‘evidentiary gap’	220
4.1 Dispelling possible misconceptions.....	220
4.1.1 <i>McGhee</i> : the defendant definitely caused the loss.....	221
4.1.2 <i>Fairchild</i> : the ‘innocent claimant’ argument.....	222
4.2 Where next?.....	226
4.2.1 Mesothelioma.....	226
4.2.2 Other diseases involving an evidentiary gap	227
4.2.2.1 The single harmful agent rule: <i>Wilsher v Essex Area Health Authority</i>	228
4.2.2.2 Limiting the <i>Fairchild</i> exception to mesothelioma.....	232
Conclusion	239

Table of Cases

Allen v British Rail Engineering Ltd (BREL) [2001] EWCA Civ 242, [2001] ICR 942

Bailey v Ministry of Defence and another [2007] EWHC 2913 (QB); [2008] EWCA Civ 883, [2009] 1 WLR 1052

Baker v Willoughby [1970] AC 467 (HL)

Barker v Corus (UK) plc [2006] UKHL 20, [2006] 2 AC 572

Barnett v Chelsea and Kensington Hospital Management Committee [1968] 2 WLR 422 (QBD)

Bolitho v City of Hackney Health Authority [1998] AC 232 (HL)

Bonnington Castings v Wardlaw [1956] AC 613 (HL)

Cartledge v E Jopling & Sons Ltd [1963] AC 758 (HL)

Chaplin v Hicks [1911] 2 KB 786 (CA)

Chester v Afshar [2004] UKHL 41, [2005] 1 AC 134

Dean v Keate (1811) 3 Camp. 4 (170 ER 1286)

Dickins v O2 plc [2008] EWCA Civ 1144, [2009] IRLR 58

Dillon v Twin State Gas & Electric Co, 85 NH 449, 169 A 111 (1932)

Dingley v The Chief Constable, Strathclyde Police 1998 SC 548

Durham v BAI (Employers' Liability Insurance "Trigger" Litigation) [2012] UKSC 14, [2012] 1 WLR 867

Jobling v Associated Dairies [1982] AC 794 (HL)

Fairchild v Glenhaven Funeral Services (t/a GH Dovener & Son) [2001] EWCA Civ 1881, [2002] 1 WLR 1052; [2002] UKHL 22, [2003] 1 AC 32

Gouldsmith v Mid-Staffordshire General Hospitals Trust [2007] EWCA Civ 397, [2007] LS Law Medical 363

Gregg v Scott [2002] EWCA Civ 1471, [2003] Lloyd's Rep Med 105; [2005] UKHL 2, [2005] 2 AC 176

Gwilliam v West Herts NHS Trust [2002] EWCA Civ 1041, [2003] QB 443

Hatton v Sutherland [2002] EWCA Civ 76, [2002] 2 All ER 1

Hinze v Berry [1970] 2 QB 40 (CA)

Holby v Brigham & Cowan (Hull) Ltd [2000] 3 All ER 421 (CA)

Hotson v Fitzgerald and others [1985] 1 WLR 1036 (QBD)

Hotson v East Berkshire Health Authority [1987] 2 WLR 287 (CA); [1987] AC 750 (HL)

Jobling v Associated Dairies Ltd [1982] AC 794 (HL)

Lim Poh Choo v Camden & Islington AHA [1980] AC 174 (HL)

McGhee v National Coal Board 1972 SLT (notes) 61 (Court of Session, Scotland); [1973] 1 WLR 1 (HL (Sc.))

McWilliams v Sir William Arrol Co Ltd [1962] 1 WLR 295 (HL)

Nicholson v Atlas Steel Foundry & Engineering Co Ltd [1957] 1 WLR 613 (HL (Sc))

Overseas Tankship (UK) Ltd v Morts Dock and Engineering Co Ltd (The Wagon Mound (No1)) [1961] AC 388 (PC)

Performance Cars v Abramam [1962] 1 QB 33 (CA)

Rahman v Arearose Ltd [2001] QB 351 (CA)

Re Polemis and Furness Withy Co [1921] 3 KB 560

Reeves v Commissioner of Police of the Metropolis [2000] 1 AC 360 (HL)

Rothwell v Chemical and Insulating Co Ltd and another [2007] UKHL 39, [2007] 3 WLR 876

Sienkiewicz v Greif (UK) Ltd [2011] UKSC 10; [2011] 2 WLR 523

Sindell v Abbott Laboratories 26 Cal. 3d 588, 607 P 2d 924 (1980)

Smith v Leech Brain & Co [1962] 2 QB 405 (QBD)

South Australia Asset Management Corp (“SAAMCo”) v York Montague Ltd [1997] AC 191 (HL)

Thompson v Smiths Shiprepairs (North Shields) Ltd [1984] QB 405 (QBD)

Williams v University of Birmingham [2011] EWCA Civ 1242, [2012] PIQR P4

Wilsher v Essex AHA [1987] 1 QB 730 (CA); [1988] AC 1074 (HL)

Introduction

It has been said that ‘[t]o insist on a causal connection between conduct ensures that in general we impose liability only on those who, by intervening in the world, have changed the course of events for the worse’.¹ The factual causation requirement is therefore an essential ingredient of negligence liability, yet in recent years the approaches of courts and academics to the doctrine of causation have become increasingly complicated and confused.

The primary aim of this thesis is to identify a coherent legal response to the particular causal problem of the ‘evidentiary gap’. The ‘evidentiary gap’ problem arises where a claimant has been exposed to a harmful agent from a number of sources, including the defendant’s negligence, and a gap in scientific understanding of the aetiology of the disease prevents proof that any particular source of the harmful agent was actually a cause of the harm. This is a rare problem that arises because medical science cannot explain how the disease occurs so it represents an insurmountable problem of proof for the claimant. This was first addressed in *McGhee*² and *Fairchild*,³ where the House of Lords relaxed the causation requirement and allowed the claimants to recover on the basis of proof that the defendant’s negligence had ‘materially increased the risk of harm’. In order to identify a coherent solution to the ‘evidentiary gap’ it will first be necessary to situate the issue within the broader context of causation in negligence more generally. This will involve addressing both the nature and function of the tort of negligence as well as the role played by causation within that tort.

The thesis will therefore begin with an analysis of the theoretical basis of negligence. In Chapter One, it will be argued that negligence is best understood in terms of corrective justice-based interpersonal responsibility because this provides an account of negligent that prioritises coherence and morality in the law. After explaining that causation occupies a pivotal yet limited

¹ Tony Honoré, *Responsibility and Fault* (Hart Publishing 1999) 120.

² *McGhee v National Coal Board* [1973] 1 WLR 1 (HL (Sc.)).

³ *Fairchild v Glenhaven Funeral Services* (t/a GH Dovener & Son) [2002] UKHL 22, [2003] 1 AC 32.

role within interpersonal responsibility, the chapter will turn to the doctrinal framework of negligence in order to show how this role for causation is to be translated into practice. This section will illustrate the interrelationship between the various negligence doctrines which each fulfil a distinct purpose yet join together to form an integrated whole. Subsequent analysis of caselaw on causation will draw on this to show that some problems which are currently treated as being causal are not truly causal and that they should be addressed through other negligence doctrines. The analysis of the doctrinal framework will also highlight the centrality of the duty concept, and will argue that the issues of interpersonal responsibility identified at the duty of care stage of the negligence inquiry should shape the application of the remaining doctrines. Later chapters will draw on this idea when addressing problems of causation.

Chapter Two then turns to the concept of causation and the demands that the law places on a test for causation. Since causation is pivotal to interpersonal responsibility it is important that the legal approach to causation is premised on a philosophically sound account of the concept of causation. In this chapter it will be argued that Wright's NESS test more accurately translates the philosophical account of causation into a workable test than the but-for test that currently dominates judicial approaches to causation. In the majority of negligence cases where causation is relatively straightforward, the NESS test is no more complicated to apply than the but-for test, but in the small number of cases where the causal problem is more complex the NESS test is able to identify causes where the but-for test fails.

It will be demonstrated that, under NESS, it is essential to define the damage that forms the gist of the negligence action, and that the causal problem will vary depending on whether the harmful outcome is 'divisible' i.e. dose-related, or 'indivisible' ('all-or-nothing'). Chapter Two will also evaluate the *Wardlaw* test of 'material contribution to harm'.⁴ It is currently unclear when and why this test applies, and therefore it is unclear whether it is an exception to the but-for test and/or to the causation requirement. Using the NESS analysis it will be shown that the *Wardlaw*

⁴ *Bonnington Castings v Wardlaw* [1956] AC 613 (HL).

test seems to address a wide range of causal problems: where damage is divisible it equates to an application of the but-for test to a portion of the harm, where damage is indivisible it is often an exception to the but-for test but not an exception to the causation requirement. In these circumstances the *Wardlaw* test compensates for the conceptual inadequacies of the but-for test. This has important consequences for the approach to evidentiary uncertainty. The NESS test will enable us to see that cases where the *Wardlaw* test is applied are not really problematic, nor are they exceptional. Since the *Wardlaw* test is not exceptional, it cannot be used as a stepping-stone towards the adoption of exceptional solutions to problems of evidential uncertainty as happened in *McGhee*.

Having addressed the conceptual aspects of causation in Chapter Two, the remaining chapters turn towards evidential problems relating to causation. Chapter Three considers the so-called ‘loss of a chance’ argument that was raised by the claimants in *Hotson*⁵ and in *Gregg*.⁶ These claimants arose in the context of medical negligence resulting in a delay in the diagnosis and treatment of patients who were already unwell. Since the patient’s prospects were poor at the time of the negligence, they faced difficulty in proving that the negligent delay in diagnosis had caused them to suffer physical harm. Instead they sought to redefine the damage forming the gist of the action as the loss of a chance of avoiding physical harm. Although this argument centres on the definition of the relevant damage, its inclusion in this thesis is important for a number of reasons. The claimant’s attempt to redefine the damage was motivated because of difficulties of proving a causal link between the negligence and the physical harm, and one of the recurring issues addressed in this thesis is the relationship between the doctrines of damage and causation. It is also essential to understand and evaluate the loss of a chance argument because ‘loss of a chance of avoiding harm’ is often considered to be the equivalent of ‘a material increase in the risk of harm’ which is contained in the *McGhee/Fairchild* exception in cases of evidential uncertainty.

⁵ *Hotson v East Berkshire Health Authority* [1987] AC 750 (HL).

⁶ *Gregg v Scott* [2005] UKHL 2, [2005] 2 AC 176.

In Chapter Four, the discussion turns to an analysis of the approaches taken by courts and academics to the problem of evidentiary uncertainty in causation, drawing together the themes running through the previous chapters. It begins by evaluating the evidence relating to the diseases in *McGhee* and in *Fairchild* in order to understand more precisely what the ‘evidentiary gap’ consists of. The NESS account of causation developed in chapter two enables the ‘evidentiary gap’ to be defined with greater clarity than under the but-for test. The chapter then engages with the *McGhee/Fairchild* test of ‘material increase in the risk of harm’ as a test of sufficient causal connection.

Following *McGhee* it was suggested that the court had simply taken a ‘robust’ view of the available evidence, drawing an inference that the negligence had materially contributed to the harm from the fact that it had materially increased the risk of that harm. It will be argued that the nature of the evidentiary gap means that it is simply not possible to draw a rational inference of causation, so this interpretation rests on a fiction.

Following *Fairchild*, the court in *Barker* apportioned liability according to the defendant’s contribution to the total risk to which the claimant was exposed, and Lord Hoffmann rationalised this on the basis that because the claimant was only able to prove that the defendant had materially increased the risk of harm, it was the risk of harm rather than the harm itself that constituted the gist of the negligence action.⁷ Chapter Four analyses the notion of risk in order to determine whether risk can, and did, form the gist of the negligence action in that case. Ultimately it will be argued that if risk exposure was the damage forming the gist of the action, the damage requirement would be subsumed into the breach inquiry which asks whether the defendant exposed the claimant to an unreasonable risk of harm. This means that liability would be based solely on the careless conduct of the defendant so liability would be based on a retributive form of justice rather than on corrective justice.

⁷ *Barker v Corus (UK) plc* [2006] UKHL 20, [2006] 2 AC 572, [35] (Lord Hoffmann).

Concluding that a causal link between the negligence and the physical harm cannot be proved in cases involving an evidentiary gap, the chapter will finally consider the arguments that liability should nonetheless be imposed because of policy arguments relating to the perceived need to compensate the victims of mesothelioma. This section will draw on the exposition of the theoretical foundations of negligence from Chapter One in order to evaluate the policy concerns and arguments as to the ‘demands of justice’.

Therefore while this thesis takes the analysis of the approaches taken to evidential uncertainty as its ultimate objective, its contribution to the study of causation extends beyond that narrow problem. This means that the thesis achieves more than the title might imply, contributing to the understanding of the basic tests of causation and seeking to add clarity to the relationship between the causation requirement and the other negligence doctrines, especially the notions of damage and quantification. It also means that the arguments that are made in relation to the evidentiary gap are part of an overarching vision of negligence liability and the place of causation within negligence law, so it makes insights into evidentiary gap that might otherwise be lost, and gives these arguments a much stronger foundation. Much of the confusion surrounding causation arises from courts and academics discussing problems of causation too much in isolation and without giving sufficient weight to the theoretical and doctrinal framework of the negligence enquiry. This thesis seeks to show that while causation is a complex field, it need not be confusing.

Chapter One: Theoretical and doctrinal framework

Introduction

In order to resolve problematic causal issues in a way that is coherent with the body of negligence law it is essential to have a clear understanding of the theoretical basis of negligence. When exceptional approaches to causation are proposed they are usually based on functionalist arguments such as the need to compensate the claimant or to achieve an economically efficient solution. It will be argued that these theories cannot provide a coherent basis for negligence so it is inappropriate to invoke them in difficult cases of causation. The purpose of the first part of this chapter is, therefore, to set out the legal theory that underlies the tort of negligence. Faced with a choice between instrumentalist theories and corrective justice-based theories it will be argued that it is corrective justice that provides the most appropriate theoretical justification for the tort of negligence. This is because the bipolar structure of the negligence action reflects the bipolarity of corrective justice and prevents the attainment of instrumentalist conceptions which are distributive, and therefore multi-polar, in nature. As such the interpersonal morality encapsulated by corrective justice ought to be reflected in the interpersonal responsibility enforced through negligence. However, the strict formalism that is characteristic of Weinrib's approach to corrective justice is arguably divorced from the practical and realistic need to take account of how liability decisions impact on wider society.¹ It would be inappropriate to insist on giving effect to interpersonal responsibility through negligence liability where this would have disproportionately adverse effects on community welfare. As Robertson explains, '[i]n the absence of adequate justification on the basis of [corrective] justice considerations, a duty cannot be imposed in the public interest, however strong the community welfare justifications'. But, as

¹ Peter Cane, 'Tort Law as Regulation' (2002) 31 CLWR 305, 310.

he adds, ‘a duty can be denied in the public interest, however strongly considerations of [corrective] justice favour the recognition of a duty’.²

The second source of strain that is placed on the doctrine of causation in exceptional cases is that the causation requirement is used to address issues that are not really causal and ought to be addressed by the other negligence doctrines. The chapter will therefore consider how the theory of corrective justice is translated into concrete legal rules through the various doctrines of the tort of negligence. It will be seen that factual causation plays a vital but limited role in corrective justice-based liability, and that it is important to reflect this in the legal rules relating to causation. Most importantly, non-causal issues relating to the concept of responsibility must be addressed through the other doctrines. Isolating the role of causation within responsibility in this way will help to clarify the causal issues in particular cases. Chapter Two will turn to the question of ‘what causation is’, but first this chapter is concerned with identifying what causation ‘is not’, in other words with identifying those aspects of responsibility that are not causal and showing how they are addressed by the other doctrines.

1. The theoretical basis of negligence

It is important to note from the outset that the focus of this thesis is on the tort of negligence so it is beyond the scope of this work to establish the theoretical foundations of tort law or private law more generally. What is sought is the theoretical foundation of the tort of negligence. England has contrasted the range of possible approaches:

On the more theoretical level: Purism, formalism, postulates of intelligibility, and coherence clash with pragmatism and pluralism; on the more ideological, political level: Individualism, right-oriented personal autonomy contrast with instrumentalism, utilitarianism, economic efficiency, and social justice.³

² Andrew Robertson, ‘On the Function of the Law of Negligence’ (unpublished conference paper, Obligations VI: Challenging Orthodoxy conference, Canada 18th-20th July 2012)

³ Izhak England, *Corrective and Distributive Justice: From Aristotle to Modern Times* (Oxford University Press 2009) 205.

Clearly the theories are many and varied. Since this thesis is focused on problems of causation within negligence, it is beyond the scope of this work to engage in a detailed analysis of all the possible theories. The objective of this chapter is therefore to signal which of the above values are most important to negligence and to identify the theoretical position which provides the best basis for the tort of negligence.

Interpretive legal theories are often evaluated against the four criteria enumerated by Smith: *fit*, *transparency*, *coherence*, and *morality*.⁴ Indeed, it has been said that '[a]lthough 'interpretivism' is sometimes treated as a particular camp within private law scholarship, Stephen Smith's statement of interpretive legal theory describes what is really an orthodox and widely followed approach to legal analysis'.⁵ It should therefore be uncontroversial to refer to these criteria to assess the theories encountered in this chapter. Robertson has helpfully summarised them as follows:

Those four limbs are: *fit* (the extent to which the theory is consistent with the outcomes of cases and possibly also the accepted rules of the given body of law), *coherence* (the extent to which the theory reveals an intelligible order in the given body of law and allows it and related bodies of law to be understood as a unified system), *transparency* (the extent to which the theory is consistent with the explanation given by lawmakers, which in the case of interpreting the common law means the explanations given by judges in the cases), and *morality* ('how the law might be thought to be justified even if it is not justified').⁶

Robertson continues to explain how these criteria are used:

The two most significant points of difference between scholars in the application of the interpretive method relate to the weight to be attributed to each of the four elements and the kinds of normative criteria that occupy the place of 'morality' in the framework...Different explanations will score better on different elements. A scholar

⁴ Stephen A. Smith, *Contract Theory* (Clarendon Press 2004) 7-32.

⁵ Andrew Robertson, 'Rights, Pluralism and the Duty of Care' in Donal Nolan and Andrew Robertson (eds) *Rights and Private Law* (Hart Publishing 2011) 438

⁶ ibid 437-8. Referring to Smith (n4).

must therefore make a choice as to which of the criteria should be weighted more heavily.⁷

The reason for preferring corrective justice over pluralist theories is that coherence ought to be prioritised and, as a monistic theory, corrective justice scores highly for coherence whereas pluralist approaches are necessarily weak in this respect. Furthermore it will be argued that a monistic theory based on an instrumentalist goal such as economic efficiency is also lacking in coherence and moral force. As Wright has explained, coherence is important because without it the law would be indeterminate:

[W]hen in a particular situation two or more of the pluralistic norms conflict – which usually will be the case – the theory will be normatively, descriptively and analytically arbitrary and indeterminate in terms of specifying which competing norm(s) should predominate, unless there is some foundational norm that can resolve conflicts between the competing subnorms. Yet if such a foundational norm exists, the theory at its deepest level is monistic rather than pluralistic.⁸

Given that coherence is valued to such an extent, it is important to understand more precisely what coherence entails. Wright continues:

This is not to say that any single norm, no matter how fundamental, can explain and justify *every* aspect of the law in general, or any particular area of law...But a successful normative and descriptive theory of law should at least be able coherently to explain and justify the principal features of the existing law.⁹

Beever, whose corrective justice-based approach to negligence prioritises coherence, draws on Smith's work to identify two senses in which a theory can be described as coherent: weak coherence requires the law to be non-contradictory, strong coherence requires that the law 'can

⁷ ibid 438.

⁸ Richard Wright, 'Right, Justice, and Tort Law' in D. Owen (ed.) *Philosophical Foundations of Tort Law* (Oxford University Press 1995) 160.

⁹ ibid.

be understood as a unified system, perhaps under a single principle'.¹⁰ This, he explains, rules out the notion of 'limited rationality' whereby one negligence doctrine such as duty of care could be based on one principle, and another negligence doctrine such as standard of care could be based on a contradictory principle. Although each doctrine may display coherence within itself, the incompatibility of the underlying principles of each doctrine would prevent the tort of negligence from being coherent in either the strong or weak senses outlined above. This is important to the scope of this thesis which seeks to understand causation itself but does not treat causation in isolation from the rest of negligence. Beever further argues that not only should the doctrines be compatible i.e. display weak coherence, it is desirable that they should be based on a single unifying principle i.e. strongly coherent:

Because unity leads to greater coherence, and hence to a theory possessing greater explanatory power, other things being equal, a theory that provides a unified explanation of the law in the sense elucidated is preferable to one that does not. Unity may not be mandatory, but it is attractive.¹¹

To summarise the arguments so far, coherence is desirable because of the justificatory force it carries and because it avoids the indeterminacy inherent in pluralist approaches. It is desirable for a theory to display 'strong coherence' in the sense of enabling the law to be understood as a unified system, which means that the theory must be able to explain the principal features of the law even if it is unable to explain every detail of the law.

As mentioned above, Wright correctly asserts that '[o]nly a monistic foundational theory holds out any prospect of being able to do [this]',¹² and he goes on to explain that there are two principal monistic theories: '(1) utilitarian efficiency theory, based on the foundational norm of

¹⁰ Allan Beever, *Rediscovering the Law of Negligence* (Hart Publishing 2009) 22.

¹¹ ibid 23-24.

¹² Wright, 'Right, Justice and Tort Law' (n8) 160.

maximizing aggregate social welfare...and (2) the Kantian-Aristotelian theory of Right or justice, based on the foundational norm of equal individual freedom'.¹³

It is the Kantian-Aristotelian theory of justice that provides the more coherent account of negligence, so the focus of this section of the chapter is to illustrate what the theory consists of and why it provides the most coherent justification of the tort of negligence. Later sections will then consider the implementation of this theory through the negligence system and through the doctrines of negligence law.

1.1 The Aristotelian conception of corrective justice

The starting point to a corrective justice-based account of negligence must be to explain the Aristotelian conception of corrective justice. It is clear that Aristotle's account of corrective justice is abstract and is not an account of the law, let alone an account of a modern tort system, so it is essential to understand the crucial features of this abstract account that must be translated into the practices of negligence law in order for the negligence system to be consistent with corrective justice.

1.1.1 Justice as a whole

Corrective justice, along with distributive justice, was originally elaborated by Aristotle in the *Nicomachean Ethics*.¹⁴ Aristotle's discussion of justice follows his exposition of the 'excellences' of character which are intermediate qualities between having too much and too little of contrasting defects.¹⁵ Similarly, in Aristotle's account, justice must also be an intermediate state. Justice, for Aristotle, exists as a whole and as a part. Justice as a part is a quantitative state and consists in 'equal-mindedness' as opposed to 'graspingness' in relation to goods. As Gardner explains,

¹³ ibid.

¹⁴ Broadie & Rowe, *Aristotle: Nicomachean Ethics: Translation, Introduction and Commentary* (Oxford University Press, Oxford 2002) (cited in this paper as "NE"). For ease of reference, citations will be made to chapter and line numbers from Aristotle's original text rather than to pages of this particular translation.

¹⁵ NE Books II-IV. For example, courage is an excellence which is the intermediate between fear and boldness.

‘Norms of justice are moral norms of a distinctive type. They are norms for tackling *allocative* moral questions, questions about who is to get how much of what’.¹⁶ Justice as a whole is a qualitative concept which comprises justice as a part along with the quality of being a law-abiding person.¹⁷ The law, if it has been laid down correctly, requires a person to display the other excellences of character (such as courage, moderation, mildness) towards other people. Justice, in both forms, is an excellence which is a state relative to another individual rather than being an internal disposition: ‘This justice, then, is complete excellence, only not without qualification but in relation to another person’.¹⁸ Aristotle, having already explored the other excellences of character in depth, therefore turns to ‘justice as a part’ to address it more fully and it is here that we are introduced to the subdivision of corrective¹⁹ and distributive justice.

1.1.2 Justice as a part: corrective and distributive justice

Justice as a part is a quantitative concept of having one’s share of goods and consists of being ‘equal-minded’:

The unjust person does not always choose the greater share, actually choosing less in the case of things that are generally bad; but since the lesser evil too is regarded as good in a way, and what the grasping person wants is more of what is good, this makes people think of him as grasping.²⁰

It is therefore the intermediate state between having too much and too little both of what is good and of what is bad. And since justice is ‘other-regarding’ it is relational and consists in having one’s share compared to another person. Aristotle explains that there are two aspects to having one’s share:

¹⁶ John Gardner, ‘What is Tort Law For? Part 1: The Place of Corrective Justice’ (2011) 30 *Law and Philosophy* 1, 6.

¹⁷ *NE* V.1 (1129a32-34).

¹⁸ *NE* V.1 (1129b26-27).

¹⁹ Referred to by Aristotle as ‘rectificatory’ and later by Thomas Aquinas as ‘commutative’.

²⁰ *NE* V.1 (1129b6-10).

Of the justice that is a part, and of what is just in this sense, one sort is the one found in distributions of honour, or money, or the other things to be divided up among those who are members of the political association (for in the case of these things it is possible for one person to have either an unequal or an equal share in relation to another); while another is rectificatory, operating in interactions between one person and another.²¹

The first sort, distributive justice, determines what is just within a community and is an equality of ‘geometrical’ proportion, while corrective justice determines what is just in an interaction between two individuals and is an equality of ‘arithmetical’ proportion. Although corrective justice is the form that corresponds to negligence since it addresses justice in interactions, it can only be fully understood in relation to distributive justice so both forms will be explained. This is particularly important since consequentialist justifications of modern negligence law, such as the utilitarian imperative of maximising wealth which is seen in economic efficiency theories, are concerned with consequences that affect the community rather than merely the two parties to the interaction. They are therefore distributive justice-based approaches to negligence so it is important to understand what distributive justice is, and to illustrate the impossibility of achieving distributive justice through an institution, such as negligence law, that takes the structural form of corrective justice.

1.1.2.1 Distributive justice

Distributive justice addresses the distribution of goods among the members of a community. It is an equality of geometrical proportion, so it does not require every individual to have an equal share; rather it involves a distribution according to merit. Thus it requires that an individual’s share of goods (compared to others’ shares) is equal to his level of merit (compared to the merit of others). Distributive justice does not entail any particular criterion of merit:

²¹ NE V.2 1130b30-1131a2.

[E]verybody agrees that what is just in distributions must accord with some kind of merit, but everybody is not talking about the same kind of merit: for democrats merit lies in being born a free person, for oligarchs in wealth or, for some of them, in noble descent, for aristocrats in excellence.²²

To use a common analogy, distributive justice is akin to sharing a cake and determining how big a slice of cake each person should receive. There is a limited pot of resources since the cake is finite, and it must be shared according to some criterion of merit, perhaps according to who contributed the most ingredients or work, who is the most hungry or needy, or all persons equally. The important point is that for limited resources to be shared in a way that is distributively just there must be some criterion of merit and a person's share should be equal to his worth according to that particular measure of merit. Distributive justice is not, therefore, an appeal to what the man on the Underground might consider fair.²³ It is a comparative exercise, so is closer in nature to Lord Hoffmann's concern in *White* 'to preserve the general perception of the law as system of rules which is fair *between one citizen and another*',²⁴ although 'fairness' is inevitably still dependent on what criteria of distribution is prioritised.

1.1.2.2 Corrective justice

In contrast, corrective justice concerns the meaning of 'having one's share' in an interaction between two individuals. Where distributive justice involved equality of geometric proportion, corrective justice involves an equality of arithmetical proportion.²⁵ This arithmetical proportion is derived from the pre-transaction equality of the parties:

For it makes no difference whether a decent person has defrauded a worthless one or a worthless person a decent one...the law pays attention solely to the difference created by

²² NE V.3 1131a25-30.

²³ Cf *White v Chief Constable of South Yorkshire* [1999] 2 AC 455 (HL), 495 (Lord Steyn).

²⁴ ibid 511 (emphasis added).

²⁵ NE V.4 1131b25-1132a2.

the damage done, and where one person is committing an injustice, another suffering it, or one person inflicted damage and another has been damaged, it treats them as equal.²⁶

This notion of equality will be revisited in more depth later, but at this stage it is sufficient to note that Aristotle regards the parties as equal irrespective of whether they have equal holdings, irrespective also of whether they are assessed as equal according to any particular criterion of merit, and irrespective of whether their pre-transaction holdings were distributively just. As Weinrib has explained, ‘The parties do not have the same quantity of holdings, but they are equal as the owners of whatever they do have’.²⁷

When one person violates this equality by injuring another he is considered to make a gain which is equal to the other’s loss, so equality can be restored by requiring him to repair the other’s loss. It is clear that in concrete terms the gain made by an injurer is not the mirror image of the loss he inflicts. Aristotle therefore says it may be ‘inappropriate’ to talk of a gain – in layman’s terms there has been no material gain.²⁸ If I break somebody’s arm I do not ‘gain’ an arm myself. But because a gain consists not only in having too much of what is good, but also in having too little of what is bad, I do make a gain through indulging my careless behaviour and through my own lack of suffering. The gain exists as a violation of the equality of the rights of the parties.²⁹

Because both parties to an interaction are considered equal prior to the interaction, this gain is correlative to the loss inflicted on the victim. The pre-transaction equality was not an equality of holdings, nor an equality of status, so a gain is not measured in terms of the concrete increase in goods or status. The pre-transaction equality was an abstract equality of right to be free from

²⁶ NE V.4 1132a3-1132a7.

²⁷ Ernest Weinrib, ‘Corrective Justice’ (1992) 77 Iowa L Rev 403, 408.

²⁸ NE V.4 1132a10-1132a19.

²⁹ Gordley draws on the work of Thomas Aquinas to explain this: “Thomas clarified: a “person striking or killing has more of what is evaluated as good, insofar, that is, as he fulfils his will, and so is seen to receive a sort of gain”. To gain, then, means to fulfil one’s will. One who has taken or used or harmed another’s resources for his own ends has ‘gained’, and therefore must pay compensation, whether or not his ends have been achieved, and whether or not he has made a financial gain by pursuing them” (James Gordley, ‘Tort Law in the Aristotelian Tradition’ in David G Owen (ed.) *Philosophical Foundations of Tort Law* (Oxford University Press 1995) 138. Citing Aquinas, *In decem libros ethicorum Aristoteles exposition*, (Angeli Pirotta (ed.) 1934), lib. V lectio vi no. 952)).

injury, so the extent of the injury inflicted on another represents the extent to which the wrongdoer has taken more than he should. As Gordley has explained:

By voluntarily harming the plaintiff, [the defendant] has chosen to use the plaintiff's resources for his own ends. The pre-existing equality that corrective justice seeks to restore is a state in which each party achieves his own goals out of his own resources.³⁰

Again, the concept of equality will be revisited in more detail, but for now it is a sufficient explanation of the basic Aristotelian account of corrective justice in so far as the key features and structures of each form of justice have been set out clearly.

In practical terms the correction is made by requiring the injurer to repair the victim's loss and this is achieved through compensation because, as Aristotle explains, currency 'acts like a measure, making things commensurable'.³¹

In summary, the key features of corrective justice are that it addresses justice in interactions so it is bipolar, the two parties are treated as equal before the interaction which gives rise to a correlative gain and loss and necessitates a correction which is also correlative i.e. the injustice can only be corrected by requiring the wrongdoer to repair the victim's loss. As a form of justice it is distinguished from distributive justice which is multi-polar since it addresses justice across a community, which achieves a geometrical equality of holdings by specifically recognising the inequality of individuals according to a criterion of merit, and which necessitates a broad institution to implement the correct distribution across the community.³²

³⁰ *ibid* 138.

³¹ *NE* V.5 1133b16

³² It is important to signal these key features because not all 'corrective justice' theories of negligence actually reflect these characteristics. One of the primary problems for Coleman's 'annulment' theory was that it was not based on correlativity, and this led to him abandoning it in favour of a 'mixed conception' of corrective justice which also incorporates a relational approach. Coleman identified his annulment theory as being one of corrective justice because, for him, corrective justice required the annulment of wrongful gains and losses. The interpersonal character of corrective justice was not reflected, however, in Coleman's account where the victim's claim for compensation was based on wrongful loss and the imposition of liability on the defendant was based on wrongful gain, so the grounds of recovery were conceptually distinct from the grounds of liability. This violated the correlative character of corrective justice where the grounds of recovery are the same as the grounds of liability i.e. the loss and gain derive their 'wrongful' character from the same source. In his 'mixed conception' he incorporates a relational approach, explaining that the relational conception of 'wrong' explains why the wrongdoer and nobody else has a duty to repair the loss. His work will not be addressed in depth, however, because his approach is still not based on a single overarching principle as required for strong coherence since he takes a relational approach to

1.2 The moral basis of corrective justice: Aristotelian corrective justice and Kantian Right

The formalistic reading of the Aristotelian conception of corrective justice has given an insight into the distinctive nature of corrective justice and the demands of coherence in terms of maintaining a separation between corrective and distributive justice. But this account is incomplete insofar as the concept of equality has not been fully explained, so the content and moral basis of corrective justice remains obscure. It is vital to understand that corrective justice is not merely a procedural requirement; it is not ‘an empty formalist shell’.³³ Beever argues:

According to corrective justice theory, while corrective justice tells us how the injustice should be corrected, it must also be true that the injustice itself is interpersonal. That is, the structure of corrective justice must be reflected in the nature of the wrongdoing to which corrective justice responds.³⁴

This means that corrective justice cannot accommodate an account of wrongdoing that is distributive, such as aggregate social welfare. Posner has proposed such an approach, arguing that corrective justice dictates form but not substance so the concept of wrongdoing must be sought outside Aristotle’s corrective justice and can be based on aggregate social welfare.³⁵ The formalist reading of corrective justice showed, however, that this would lead to incoherence because liability would be based on a distributive premise. Strong coherence requires that the principle of corrective justice is reflected not only in the structure but also in the rules of negligence liability. Englard has explained, ‘it has to be remembered that even from a strictly formalist conception, the difference in the formal structures implies some important

explain the wrong, and an annulment approach to explain the loss. Jules Coleman, ‘Corrective Justice and Wrongful Gain’ (1982) 11 J Legal Stud 421; Jules Coleman, ‘The Mixed Conception of Corrective Justice’ (1992) 77 Iowa L Rev 427.

³³ Wright, ‘Right, Justice and Tort Law’ (n8) 170. Wright also criticises the ‘formalist evisceration’ of corrective justice: Richard Wright, ‘Substantive Corrective Justice’ (1992) 77 Iowa L Rev 625, 710.

³⁴ Allan Beever, ‘Corrective Justice and Personal Responsibility in Tort Law’ (2008) OJLS 475, 477.

³⁵ Richard Posner, ‘The Concept of Corrective Justice in Recent Theories of Tort Law’ (1981) 10 J Legal Stud 187, 191-201.

consequences for the substances. Form and substance become linked through the idea of coherence'.³⁶

In contrast, Kantian Right, espoused by academics such as Weinrib and Wright, provides an understanding of wrongdoing that conforms to the requirement of correlativity. As Weinrib states, 'the Kantian interpretation fits readily into, and provides content for, corrective justice'.³⁷

Although the ideas of Aristotelian corrective justice and Kantian Right are not explicitly articulated by courts, 'they are implicit in [private law] as a coherent justificatory enterprise, in that they provide its unifying structure and its normative idea'.³⁸ It is Kantian Right that provides the moral theory, or normative idea, underlying negligence.

In combining corrective justice and Kantian Right, Weinrib begins with Aristotle's statement that 'the law treats [the parties to the interaction] as equals however much they may be unequal' and discerns three distinct features of corrective justice: first corrective justice treats the parties as equals; second corrective justice takes no account of any measure of merit of the parties; and third it focuses on the 'immediate relationship of doer to sufferer'.³⁹ Weinrib explains that corrective justice therefore brings together the three ideas of 'the equality of the parties, the abstraction from such particulars as social status and moral character, and the conception of agency present in immediate interaction'.⁴⁰ The conception of self-determining agency which is inherent in Kantian Right also displays these three qualities:

The fundamental feature of this agency is the agent's capacity to abstract from – and thus not to be determined by – the particular circumstances of his or her situation. Inasmuch as this capacity is a defining feature of self-determining agency, all self-determining agents are equal with respect to it.⁴¹

³⁶ England (n3) 194.

³⁷ Weinrib, *The Idea of Private Law* (Harvard University Press 1995) 19.

³⁸ *ibid* 19.

³⁹ Weinrib, 'Corrective Justice' (n27) 421.

⁴⁰ *ibid* 421-22.

⁴¹ *ibid* 422.

This leads him to conclude that Kantian Right ‘dovetails’ with Aristotelian corrective justice.⁴²

So what is Kantian Right, and how does it provide the normative basis for corrective justice?

1.2.1 Kantian Right

Kant’s theory belongs in the tradition of natural right and is based on the idea of free will. Free will, or freedom, is a characteristic possessed by every rational being and gives every rational being absolute and equal moral worth:

[M]an regarded as a person, that is, as the subject of a morally practical reason, is exalted above any price; for as a person (*homo noumenon*) he is not to be valued merely as a means to the ends of others or even to his own ends, but as an end in himself, that is, he possesses a *dignity* (an absolute inner worth) by which he exacts *respect* for himself from all other rational beings in the world. He can measure himself with every other being of this kind and value himself on a footing of equality with them.⁴³

Respect for the absolute and equal moral worth of everyone gives rise to the ‘categorical imperative’: ‘act only according to that maxim by which you can, at the same time, will that it should become a universal law’.⁴⁴ This is similar to, but wider than, the Golden Rule that you should treat others as you wish to be treated yourself:

It is morally wrong under the categorical imperative to fail to respect the absolute moral worth of anyone, including yourself, as a self-legislating rational being, regardless of whether you would allow others to treat you without proper respect.⁴⁵

Whereas the Golden Rule allows you to treat others badly if you would be happy to be similarly treated yourself, the categorical imperative does not allow this because it demands respect for one’s own moral worth. As Wright explains, ‘[t]he idea of freedom does not imply completely

⁴² ibid 422.

⁴³ Immanuel Kant, *The Metaphysics of Morals* (Mary Gregor trans., 1991) (1797) cited in Wright (n8) 162.

⁴⁴ Kant, *Foundations of the Metaphysics of Morals* (trans. Lewis White Beck) (Bobbs-Merrill Education Publishing 1969) 44.

⁴⁵ Wright, ‘Right, Justice and Tort Law’ (n8) 162-63.

unrestricted self-determination, but rather self-legislation: self-determination according with universal law'.⁴⁶ The exercise of free will in Kant's theory has both an internal aspect in the doctrine of virtue, and an external aspect in the doctrine of Right:

The supreme principle of Right is 'so act externally that the free use of your choice can coexist with the freedom of everyone in accordance with a universal law', while the supreme principle of virtue is '[a]ct in accordance with a maxim of *ends* that it can be a universal law for everyone to have'.⁴⁷

So a person who displays virtue desires ends that respect the equal moral worth of all beings and these ends can therefore be adopted by anyone. 'Right' means that even if your intended ends are not in conformity with respect for the equal moral worth of others you still act in a way that does respect their equal moral worth.

Therefore Right for Kant, just like justice for Aristotle, is directed towards others. In other words Right is concerned with the interaction with others and can be enforced through law, virtue is concerned with the reason for your choice of conduct in this interaction and is thus internal and cannot be coerced. A virtuous person would act in accordance with universal laws because he chooses ends that are consistent with those universal laws, but for Right all that matters is that the person complies regardless of the ends he is pursuing through this compliance. This seems similar to the distinction between justice as a whole and justice as a part in Aristotle's account of justice. The concern of justice as a part is that a person gives others their due, the motivation behind giving a person his due is not relevant, whereas justice as a whole is qualitative and requires one to display the other excellences of character towards others. This concept of equality which is clearly distinct from quantitative assessments also clarifies the correlativity of the gain and loss in corrective justice:

The relevant loss is the damage suffered by the plaintiff as a result of the defendant's failure to give him equal consideration...The defendant's gain is that he did not carry the

⁴⁶ ibid 162.

⁴⁷ ibid 162. Citing Kant (n43) *231, 395.

burden of appropriate precautions as he implicated the plaintiff into the web of his own projects. And the notion of equality against which these gains and losses are measured is the Kantian prohibition against self-preference in action.⁴⁸

Kantian Right has important implications for corrective justice because it provides an account of equality which is divorced from any assessment of the relative merits of the parties according to a criterion of merit, as well as being independent of the relative holdings of the parties prior to their interaction.

It therefore illuminates some of the features of corrective justice. In terms of wrongdoing, Wright says:

Unlike distributive justice, corrective justice does not use interpersonal comparisons or rankings to implement a relative or proportional equality among the parties to the interaction...this means that all comparative criteria for determining unjust gains and losses in corrective justice – including utilitarian, efficiency, and all other aggregative criteria – are excluded from consideration. This is a powerful substantive implication from Aristotle's account.⁴⁹

This means that the concept of wrongdoing in corrective justice cannot be based on economic efficiency, because this would be inconsistent with respect for the absolute moral worth of rational beings since it treats them as means to an end rather than as ends in themselves. Economic efficiency cannot be accommodated within corrective justice because ‘moral theories based on equal individual freedom give an absolute and primary status, rather than a contingent and derivative status, to the autonomy and rights of individuals’.⁵⁰

So Aristotelian corrective justice and Kantian Right together imply not only the form of the correction but also the nature of wrongdoing at its core.

⁴⁸ Ernest Weinrib, ‘Toward a Moral Theory of Negligence Law’ (1983) 2 Law & Phil 37, 53.

⁴⁹ Wright, ‘Substantive Corrective Justice’ (n33) 700-701.

⁵⁰ Wright, ‘Right, Justice and Tort Law’ (n8) 162.

1.3 Incoherence of consequentialist theories

In contrast to corrective justice, which provides a coherent account of the form of negligence law and of the wrongdoing at its core, consequentialist accounts are unable to provide a coherent theoretical foundation. Since they are distributive in nature they cannot be achieved through the bipolar structure of a negligence action. Moreover, pluralist consequentialist theories, such as the idea that negligence pursues the twin goals of compensation of victims and deterrence of wrongdoers, suffer from ‘a surfeit of reasons and norms’,⁵¹ which pose artificial limits on one another. This means that if negligence doctrines are developed in pursuit of consequentialist goals the law of negligence will become incoherent. Subsequent criticism of negligence law for being ineffective or inefficient as a system of compensation or deterrence stems from the fact that these are distributive goals so are necessarily precluded from being effectively achieved through the structure of negligence.

Wright explains that compensation and deterrence cannot provide a sound basis for negligence law because ‘compensation and deterrence of all losses is normatively insupportable, descriptively implausible, and analytically impossible’.⁵² He goes on to explain that it is normatively insupportable because there is ‘no plausible moral argument for requiring others to compensate every person for every loss no matter how it occurred’.⁵³ It is also impossible because compensation does not eradicate loss but shifts the loss to others, and because it is neither possible nor desirable to deter all risky conduct. The bipolar structure of negligence liability prevents each goal from being fully attained:

The fundamental reason why these criteria are inconsistent with tort law’s correlative structure is that there is no reason to limit the search for the deepest pocket, or the best

⁵¹ ibid 160.

⁵² ibid 159.

⁵³ ibid 159.

loss-spreader, or the cheapest cost-avoider, to the two parties to the tort claim – the harm-doer and the harm-sufferer.⁵⁴

Additionally the different distributive criteria impose artificial limits on one another:

When juxtaposed within the tort relationship, compensation and deterrence are mutually truncating. What limits compensation is not the boundary to which its justificatory authority entitles it, but the competing presence of deterrence in the same legal relationship. Thus, tort law compensates victims only when damages serve the purpose of deterrence. In the same way, tort law artificially restricts deterrence by tying deterrence not to what is needed to deter wrongdoers, but to what is needed to compensate victims. In this mixing of justifications, neither goal occupies the entire area to which it applies. Accordingly, neither in fact functions as a justification.⁵⁵

Perry objects that compensation and deterrence could be combined in a way that is not mutually truncating in a system such as compulsory first-party insurance where compensation would be achieved through insurance payouts and deterrence would be achieved through the setting of premium levels.⁵⁶ It is clear, however, that this example is premised on a different institutional structure, insurance, where a harm-sufferer would be compensated independently of the insurance premium imposed on the harm-doer. The structure is distributive and matched to the distributive functions it seeks to achieve. So although Perry may be right to assert that compensation and deterrence can be combined in a way that is not mutually truncating, it remains the case that this cannot be achieved through the negligence system which links the harm-sufferer to the harm-doer in a bipolar relationship.

This means that criticisms targeted at the inefficiency of negligence law as a system of compensation and/or deterrence ought to be directed at the normative question of whether there ought to be a system of negligence liability. The significance of this choice is thrown into

⁵⁴ Peter Cane, 'Distributive Justice and Tort Law' [2001] NZ L Rev 401, 419.

⁵⁵ Ernest Weinrib, 'The Jurisprudence of Legal Formalism' (1993) 16 Harv J L & Pub Pol'y 583, 587.

⁵⁶ Stephen R Perry, 'Professor Weinrib's Formalism: The Not-So-Empty Sepulchre' (1993) 16 Harv. J. L. & Pub. Pol'y 597, 614.

sharp relief by the concerns raised by Atiyah who made an important attack on negligence law as a system of compensation in *The Damages Lottery*. The need for financial support clearly extends beyond those injured by others' wrongdoing to those suffering due to an accident or natural causes. As a system for compensation based on need, the negligence system is obviously unjust:

Of all the disabled or handicapped people in society about ten percent suffer from birth defects, about another ten percent have been injured in accidents, and the remaining eighty percent are suffering from illnesses and conditions of natural origin. Of the total number, only about one and a half percent apparently obtain any damages at all. How is this tiny group selected for preferential treatment?⁵⁷

But this is a *distributive* injustice because it is giving preferential treatment to people injured through someone else's wrongdoing in a distributive model where the criterion of merit is need. The assertion is therefore that the negligence system introduces a second, unjustified criterion of merit according a special status to injury caused by wrongdoing and this has resulted in an unjust distribution of goods. Arguably Atiyah is criticising negligence unfairly because a system that takes the form of corrective justice cannot be expected to achieve distributive justice. Atiyah's criticism is more serious though, because he argues that the rules of negligence law have been 'stretched' by the courts in order to promote compensation.⁵⁸ In other words, the courts treat negligence law as a means of achieving compensation so they are adapting the legal rules to try to achieve compensation in more cases.

As an interpretive theory, Atiyah's explanation of negligence as a system of compensation is therefore based more strongly on 'fit' and 'transparency'. As he sees it, the courts reason in terms of promoting compensation and have shaped the law to compensate more victims of harm. And the logical conclusion at which he arrives, given that compensation is a distributive model, is that the negligence system should be abandoned in favour of a fairer compensatory system. It is important that negligence law adheres to principles of corrective justice to avoid this claim.

⁵⁷ P S Atiyah, *The Damages Lottery* (Hart Publishing 1997) 143-44.

⁵⁸ ibid 32-95.

Weinrib warns that ‘a functionalist might regard causation as an indirect way of achieving market deterrence or some other extrinsic goal, an internal account treats causation as causation, that is, as a concept that represents the unidirectional sequence from action to effect’.⁵⁹ Coherence will be lost if the courts ‘stretch’ the concept of causation in order to promote compensation and deterrence.

It is important to note that proponents of corrective justice theories of negligence such as Weinrib and Beever do not generally claim that their theories are normative in the sense that the law has to have a system of corrective justice-based responsibility. Their theories are interpretive and therefore compatible with claims that the law of negligence could be abandoned and replaced with distributive systems of compensation (such as no-fault compensation or insurance) and of deterrence (such as fines or insurance premiums). Weinrib explains:

To take a modern example, the legal regime of personal injuries can be organized either correctively or distributively. Correctively, my striking you is a tort committed by me against you, and my payment to you of damages will restore the equality disturbed by my wrong. Distributively, the same incident activates a compensation scheme that shifts resources among members of a pool of contributors and recipients in accordance with a distributive criterion. From the standpoint of Aristotle’s analysis, nothing about a personal injury as such consigns it to the domain of a particular form of justice. The differentiation between the corrective and distributive justice lies not in the different subject matters to which they apply, but in the differently structured operation that each performs on a subject matter available to both.⁶⁰

Their theories do, however, have normative force in that *if* there is to be a legal system of negligence liability, in other words a bipolar system where the victim is compensated by the person whose wrongdoing caused his loss, then the system *ought* to conform to corrective justice. Cane explains, ‘it seems clear that Weinrib’s theory is meant to have normative force to the

⁵⁹ Weinrib, *The Idea of Private Law* (n37) 11-12.

⁶⁰ Weinrib, ‘Corrective Justice’ (n27) 415.

extent that in his view, if a society chooses to organize some aspect of its life in accordance with the demands of corrective justice, it should do so consistently'.⁶¹ Since the negligence system takes the form of corrective justice then the content of the legal rules that flesh it out must also reflect corrective justice, otherwise the system will become incoherent and unjustifiable.

It is therefore clear that Atiyah and Weinrib's positions can be reconciled since both insist on the impossibility of achieving distributive justice through a corrective justice structure. The difference lies in where they proceed from this basis: Weinrib argues that the law should be matched to the structure, so the rules of negligence law ought to reflect corrective justice; Atiyah argues that the structure should be matched to the law's objective, so negligence law ought to be replaced with a fairer compensation scheme. The problem with Atiyah's analysis is that it fails to measure negligence law in its own terms.

It is one thing to acknowledge, as the next section will, that negligence liability has compensatory and deterrent effects, but it is another thing to say that it has compensatory and deterrent goals and to measure its success against those goals. As a system of compensation or deterrence it has been shown that negligence will necessarily be a failure because the goals are mutually truncating and incompatible with the correlative structure of the negligence action.

1.4 Moving beyond formalism to the legal system - combining Kantian Right and societal morality.

Corrective justice has been shown to be more than just an 'empty formalist shell'; it does not simply proscribe distributive goals but actually prescribes ideas about wrongdoing based on the Kantian conception of absolute and equal moral worth of rational beings. However, Weinrib's formalist conception, in particular his assertion that 'the purpose of private law is simply to be private law',⁶² has been criticised for being too rigid in its formalism and divorced from the

⁶¹ Peter Cane, 'Corrective Justice and Correlativity in Private Law' (1996) 16 OJLS 471, 471-72. Referring to Weinrib, *The Idea of Private Law* (n37) 228.

⁶² Weinrib, *The Idea of Private Law* (n37) 21.

realities of the legal system. This is an important and valid criticism because, as already seen, coherence is not the only quality that is desirable in a justificatory theory.⁶³ The practical implementation of corrective justice through the negligence system has effects that extend beyond the parties to society more generally so whilst issues of societal morality cannot be the driving force behind negligence law they must be taken into account. While it is clear that a pluralist account cannot provide the premise of negligence law it will be argued that acknowledging, as Cane does, that the law has distributive effects this does not entail re-conceptualising the law as an instrument of distributive justice.

Cane distinguishes between treating consequences as the goals of tort law and acknowledging them as the effects of tort law:

The fact that private law is not instrumental or distributional in its explicit purposive orientation does not mean that it does not have instrumental and distributional *effects*...To the extent that the traditional view denies that private law has instrumental and distributive effects, it is wrong. But it would be wrong to think that the common law leopard has changed (or could - or is likely to – change) its responsibility spots for instrumentalist and distributional stripes.⁶⁴

This is an important distinction because it emphasises that while consequentialist theories cannot provide the premise of negligence law, it might be appropriate to limit the implementation of corrective justice where the demands of social welfare are particularly convincing. Robertson adopts this position:

In the absence of adequate justification on the basis of [corrective] justice considerations, a duty cannot be imposed in the public interest, however strong the community welfare justifications. But a duty can be denied in the public interest, however strongly considerations of [corrective] justice favour the recognition of a duty.⁶⁵

⁶³ See text to n(4).

⁶⁴ Cane, ‘Tort Law as Regulation’ (n1) 330.

⁶⁵ Robertson, ‘On the Function of the Law of Negligence’ (n2).

The practical implementation of this argument will be fleshed out in section four on the negligence doctrines; this section makes the preliminary argument that it is theoretically sound to incorporate social welfare concerns in this limited way.

Weinrib's account of corrective justice is known for its purism, yet as Morgan has noted it has been criticised as being 'so demanding that no actual legal system has ever lived up to it'.⁶⁶ Cane explains in more detail that Weinrib's focus on coherence neglects other aspects of interpretive legal theory:

An important feature of Weinrib's 'formalistic' approach is that it is concerned with 'the law in the books', and it ignores 'the law in action' – in other words, it focuses on tort *law* at the expense of what is commonly called 'the tort *system*'. As an account of the internal normative structure of tort law, Weinrib's idea of correlativity of rights and obligations (decoupled from Kantian right) seems to me to be essentially correct. On the other hand, it paints a distorted picture of the way tort law operates in practice.⁶⁷

This tension can be illuminated by returning to Smith's four criteria against which to measure interpretive legal theories: fit, transparency, coherence and morality.⁶⁸ Weinrib's account of corrective justice scores highly in terms of coherence, but it must be remembered that Weinrib's account is not entirely an interpretive one. As discussed above, Weinrib's account is normative insofar as he argues that *if* we choose to implement a system of private law then the rules of private law *ought* to conform to corrective justice. This means that it does not claim to be perfect in terms of fit and transparency. But it is important to find a balance between the normative force of the coherence of corrective justice, and the practical force of the reality of the operation of the negligence system. This reality is that liability decisions do have redistributive effects beyond the parties which may generally be considered desirable, but not at the extreme. It will be

⁶⁶ Jonathan Morgan, 'Tort, Insurance and Incoherence' 67 (2004) MLR 384, 395 referring to arguments of Ken Kress, 'Coherence and formalism' 16 (1993) Harv J L & Pub Pol'y 639.

⁶⁷ Cane, 'Tort Law as Regulation' (n1) 310.

⁶⁸ See text to (n4).

argued that the territory on which such a balance can be sought is the remaining criterion in Smith's four: morality.

The focus so far has been on an individualistic morality inherent in Kantian Right, the function of which, Cane says, 'is to protect and promote the value of human autonomy rather than other human values such as community and solidarity'.⁶⁹ Yet, as he argues elsewhere, it would be morally wrong to ignore the wider social consequences of our actions:

All tort liability is based on some notion of personal responsibility. The fact that matters of social value are taken into account in tort law does not alter this; it is simply an outworking of the straightforward moral principle that a man's responsibilities are partly a function of the fact that man is a social being.⁷⁰

So whilst coherence should be prioritised, this should not be at the expense of morality. Although Weinrib contends that distributive and corrective justice cannot be integrated into 'any overarching form',⁷¹ it will be argued that it is wrong to ignore societal morality completely, especially once we acknowledge that the practical operation of the negligence system does have effects that extend beyond the parties to the case. Whilst the focus on autonomy is consistent with the concept of Kantian Right which lies at the heart of corrective justice and ought rightly be prioritised, it is too extreme and fails to reflect man's regard for society. At some point respect for the moral worth of the individual crosses a line to become so extreme that it disproportionately impacts on the collective interests of the other members of a society.

It could be argued that Cane is simply observing that man interacts on a societal level as well as on an individual level and that this is the equivalent of Aristotle's explanation that there are two forms of justice to correspond to these two forms of interaction. But Cane is making the important point that a person does not draw a clear, distinct line between his interactions with individuals and with the community, or rather, a person does not make an absolute distinction

⁶⁹ Peter Cane, 'Rights in Private Law' in Donal Nolan and Andrew Robertson (eds) *Rights and Private Law* (Hart Publishing 2011) 56.

⁷⁰ Peter Cane, 'Justice and Justifications for Tort Liability' (1982) 2 OJLS 30, 62.

⁷¹ Weinrib, 'The Jurisprudence of Legal Formalism' (n55) 589.

between his reasons for acting in the two different contexts. Beever identifies four different ‘spheres of morality’: personal, interpersonal, societal and international.

Personal morality asks how an individual should behave in order to be a good person.

Interpersonal morality considers interactions between one individual and another. The focus here is on how persons should conduct themselves vis-à-vis one another as two individuals rather than as isolated individuals or as members of a collective. The third sphere is societal. The concern here is how to govern society and how to regulate the behaviour of individuals for the common good. The final sphere is the international. This sphere considers the impact of our actions on everyone, whether they belong to our political community or not, or the impact of our state on other states.⁷²

For Beever, negligence liability is based on corrective justice so it reflects interpersonal morality. As such, interpersonal morality is reflected in key features of negligence law, for example the objective standard of care arises because negligence is a form of interpersonal morality rather than personal so it does not matter in negligence whether there was subjective blameworthiness.⁷³ Thinking of morality as operating in different spheres can therefore help to understand the different concerns relevant to each. But it would seem strange to think that these spheres of morality actually operate in isolation from one another. Personal morality informs interpersonal morality; its equivalent in the Kantian approach is virtue which informs Right through the shared concept of conformity with universal law. There should be a similar exchange of ideas between societal and interpersonal morality.

This is an important limit to the coherence of corrective justice when it is implemented through the practices of negligence law, and it risks appearing indeterminate and therefore blurring the line between showing an awareness for avoiding extreme redistributive effects of liability and actually adopting redistribution as a goal of liability. Yet the reality is that a theory of negligence is measured not only in terms of coherence but also in terms of morality, fit and transparency.

⁷² Beever, *Rediscovering* (n10) 42.

⁷³ Beever, ‘Corrective justice and personal responsibility’ (n34) 491.

While fit and transparency may be descriptive, both coherence and morality have prescriptive force and the effort should be made to combine them. The account of negligence set out in this thesis is still based on corrective justice but takes a view of interpersonal responsibility that combines autonomous and societal moral values. It is a moral account that is premised on the absolute and equal moral worth of beings, but which shows an awareness that individual liability decisions can sometimes effect an extreme change to the distributive pattern. Whilst consequentialist accounts are distributive and thus cannot provide a coherent basis for negligence we cannot ignore the effects that liability has and should remember that man, as a societal being, will draw the line somewhere and say that his individual needs should not be prioritised over the needs of others when the effects become too significant. So while social welfare considerations cannot ground liability, exceptionally they may limit it.⁷⁴

A final clarification is needed in relation to terminology. It is important to refer to this corrective justice-based approach as reflecting ‘interpersonal responsibility’. Morgan has referred to the principles of ‘individual responsibility and corrective justice’ as distinct from collective responsibility, and Cane has said that ‘tort law is a set of rules and principles of personal responsibility’.⁷⁵ Beever equates personal responsibility with personal morality which, as explained above, is distinct from interpersonal morality and responsibility. He explains that personal responsibility would be undermined by liability insurance because it is not the defendant personally who compensates the claimant, whereas corrective justice requires only that the defendant see to it that the claimant is compensated and ‘does not imply any moral claim about the defendant per se’.⁷⁶ Although ‘interpersonal responsibility’ is a more cumbersome expression than ‘personal responsibility’ it should be used to avoid ambiguity when referring to corrective justice-based liability.

⁷⁴ Robertson ‘On the Function of the Law of Negligence’ (n2).

⁷⁵ Cane, ‘Distributive Justice’ (n54) 403.

⁷⁶ Beever, ‘Corrective Justice and Personal Responsibility’ (n34) 495.

1.5 The place of causation in corrective justice-based interpersonal responsibility

Causation has been largely unmentioned up until this point because the focus has been on the theoretical foundations of negligence generally rather than on the theory of causation. The task now is to illustrate that causation is a central feature of corrective justice-based interpersonal responsibility and that this centrality is unaffected by the argument that the distributive effects of liability ought to be taken into account in order to balance coherence and morality. As a conception of responsibility, however, causal responsibility is not synonymous with interpersonal responsibility, so causation occupies a vital but limited role which must be supplemented by ideas about wrongdoing and wrongful loss.

1.5.1 The pivotal place of causation in corrective justice

Causation is central to corrective justice because corrective justice concerns justice in interactions and without causation there is no transaction between the parties to the interaction as Weinrib explains:

The requirement of factual causation establishes the indispensable nexus between the parties by relating their rights to a transaction in which one has directly impinged upon the other. Tort law does not typically pursue wrongful conduct in the abstract. It concerns itself with such conduct only when it materializes in harm to a given person so that compensation can flow from a particular tortfeasor to his particular victim.⁷⁷

The causation requirement in negligence is therefore crucial. It is clear from the speeches in the House of Lords' decision in *Chester v Afshar*,⁷⁸ that if causation is not established liability is no longer based on corrective justice. The defendant doctor in that case breached the duty of care owed to his patient by failing to warn her of a small risk inherent in the proposed operation. This was an unavoidable risk that would have existed even if the operation itself was carried out with

⁷⁷ Weinrib, 'Toward a Moral Theory (n48) 38.

⁷⁸ *Chester v Afshar* [2004] UKHL 41, [2005] 1 AC 134.

appropriate care. Unaware of the risk, the claimant underwent the operation which was carried out without negligence but the risk materialised resulting in injury. In negligence it must be established that the defendant's breach of duty, in this case his failure to warn of the risk, was a cause of the claimant's injury. The difficulty in this case was that the claimant very honestly admitted that even if she had been warned of the risk she may still have undergone the operation after seeking a second opinion. In other words, the defendant's failure to warn her of the risk made no difference to her decision to undergo the operation so it arguably made no difference to the outcome. Although it has been suggested that a causal link can be established,⁷⁹ a substantial body of academic opinion maintains that this was a case where the defendant's negligence was not a cause of the claimant's injury.⁸⁰ Indeed the speeches in the House of Lords generally proceed on the basis that causation could not be established.⁸¹ The case is interesting at this stage because the suggestion that liability was imposed in the absence of a causal link means that it raises issues regarding the place of causation within interpersonal responsibility.

Lord Bingham found that the claimant had failed to satisfy the but-for test of causation and asked 'whether Miss Chester should be entitled to recover even though she cannot show that the negligence proved against Mr Afshar was, in any ordinary sense, a cause of her loss'.⁸² Ultimately he and Lord Hoffmann adhered to traditional corrective justice-based negligence principles and concluded that the defendant should not be held liable in the absence of a causal link.

Given the apparent lack of a causal link between the doctor's failure to warn and the patient's injury, the reasons given in the remaining judgments in favour of liability are premised on ideas

⁷⁹ Stapleton has argued that the causal link exists because if the patient had been warned of the risk she would have sought a second opinion so the operation would have taken place on a different day. The risk itself remains constant at one to two percent, so on the balance of probabilities the risk would not have materialised on the alternative occasion: Jane Stapleton, 'Occam's Razor Reveals an Orthodox Basis for *Chester v Afshar*' (2006) 122 LQR 426.

⁸⁰ See for example Sarah Green, 'Coherence of Medical Negligence Cases: A Game of Doctors and Purse' (2006) 14 Med LR 1, 4.

⁸¹ It could be argued that *Chester* is a case where the harm should be reconceptualised in order to avoid the causal problem. This idea which will be developed in later chapters where it will be argued that in very exceptional cases considerations of interpersonal responsibility require the reconceptualization of the concept of damage to include the loss of a chance. It will ultimately be argued, however, that *Chester* is an unexceptional case so there is no justification for regarding the claimant as having lost a chance, or for regarding the interference with her right to give informed consent as damage in itself.

⁸² *Chester* (n78) [9] (Lord Bingham).

that are divorced from corrective justice. Lord Steyn focused on vindicating the right of the patient to give informed consent, explaining that ‘not all rights are equally important. But a patient’s right to an appropriate warning from a surgeon when faced with surgery ought normatively to be regarded as an important right which must be given effective protection wherever possible’.⁸³ The purpose of this right is not primarily to protect the welfare of the patient, but to ensure that ‘due respect is given to the autonomy and dignity of each patient’.⁸⁴ Yet he does not note that vindication of this right would require liability whenever a doctor fails to warn a patient of the risk regardless of whether the risk materialises in harm. The damage requirement in negligence imposes an artificial limit on this vindictory imperative.

Lord Steyn also considered it important that it was the doctor’s act of operating, if not his negligence in particular, that caused the harm, but Lord Hoffmann explained that this amounts to treating the doctor as an insurer.⁸⁵ He also explained that ‘it would excuse the doctor in a case in which he had a duty to warn but the actual operation was perfectly properly performed by someone else, for example, by his registrar’.⁸⁶ If this occurred, the compensatory objective would be truncated by the bipolar structure which prevents the court from imposing liability on the ‘someone else’. Causation is clearly pivotal to corrective justice; there cannot be corrective justice-based liability in the absence of a causal link but, as Weinrib’s formalism demonstrated, distributive arguments for liability cannot be achieved within a bipolar, corrective justice-based institution.

Lord Hope objected that:

A duty was owed, the duty was breached and an injury was suffered that lay within the scope of the duty. Yet the patient to whom the duty was owed is left without a remedy.⁸⁷

⁸³ *ibid* [17] (Lord Steyn).

⁸⁴ *ibid* [18] (Lord Steyn).

⁸⁵ *ibid* [35] (Lord Hoffmann).

⁸⁶ *ibid* [35] (Lord Hoffmann).

⁸⁷ *ibid* [73] (Lord Hope).

Yet this is the nature of corrective justice. Corrective justice is not concerned with punishing wrongdoing and it is not primarily concerned with compensating loss. Corrective justice is concerned with wrongdoing that results in loss. The causation requirement has been criticised as introducing an element of ‘moral luck’ into negligence liability.⁸⁸ Howarth suggests it is ‘a kind of privilege for defendants, a privilege that allows defendants to escape liability because of coincidences for which they can take no credit’.⁸⁹ If the focus of liability was the wrongdoing of the defendant alone then there may be some force to this argument. But the focus of corrective justice is not on either of the parties alone, it is on the inequality that occurs when the wrongdoing of one person causes a loss to another person, in other words it is only concerned with equality in *transactions* and when two people interact there is only a transaction when something passes between them i.e. one causes a loss to the other.

As explained above, corrective justice theory is normative insofar as once the law adopts a system, such as negligence law, that reflects the bipolar structure of corrective justice then the legal rules ought to conform to corrective justice otherwise they become unjustified and indefensible. This discussion of the causation requirement in *Chester v Afshar* has shown that in the absence of a causal link the courts adopt justifications for liability that are based on reasons of distributive justice which are therefore incapable of providing a coherent explanation of the basis of liability, and the effects of which are artificially limited by the corrective justice structure of negligence liability. Causation is clearly indispensable to the operation of corrective justice.

1.5.2 The limited role of causation

As noted above, Lord Steyn considered it relevant in *Chester* that it was the defendant’s act of operating, if not his negligence in particular, that caused the claimant’s injury. Honoré has suggested that in these circumstances the doctor is morally responsible for the outcome of his

⁸⁸ David Howarth, ‘Many duties of care – or a duty of care? Notes from the underground’ (2006) 26 OJLS 449, 461. See also Jeremy Waldron, ‘Moments of Carelessness and Massive Loss’ in David G Owen (ed.) *Philosophical Foundations of Tort Law* (OUP 1995) 388.

⁸⁹ Howarth (n88) 461.

non-negligent act.⁹⁰ He calls this ‘outcome responsibility’.⁹¹ He argues that people of full capacity are morally responsible for the outcomes of their voluntary actions even if the outcomes are unintended, unforeseen or even unforeseeable. Outcome responsibility is not, therefore, an explanation of negligence liability and it allows people to take both credit for the positive outcomes and discredit for the negative outcomes of their actions. Instead, by showing that people are morally responsible for the outcomes of their actions, Honoré seeks to demonstrate that legal liability that is not based on moral blame (i.e. strict liability and the application of the objective standard of care to those unable to achieve that standard) is still a form of moral responsibility.

Honoré argues that outcome responsibility is ‘crucial to our identity as persons’, without it ‘we could have no continuing history or character’.⁹² But as Cane argues, if outcome responsibility was equated with causal responsibility this would be equally damaging to our personality/agency:

If people could not claim, and did not have to accept, ownership of any conduct or outcome which was affected by luck, we would lose a secure sense of ourselves and others as agents whose conduct can have effects in the world; and we would end up seeing ourselves as ‘victims of circumstance’. On the other hand, entirely ignoring the role of luck in our lives would be equally damaging to our sense of personal identity. If people could take credit for all good outcomes of their conduct no matter how lucky, and were held responsible for all bad outcomes of their conduct no matter how unlucky, we would lose the feeling of being able to act purposively and to influence events...Consequently, we can claim ownership of some of the lucky outcomes of our

⁹⁰ Tony Honoré, ‘Medical non-disclosure: causation and risk: *Chappel v Hart*’ (1999) 7 Torts LJ 1.

⁹¹ See generally Tony Honoré, ‘Responsibility and Luck’ in Tony Honoré, *Responsibility and Fault* (Hart Publishing 1999).

⁹² *ibid* 29.

conduct, but not all; and we must accept ownership of some of the unlucky outcomes of our conduct, but need not accept all of them.⁹³

The problem with outcome responsibility is therefore that it contains an unarticulated limit on the consequences that will be considered to be outcomes of a person's conduct. Honoré's outcome responsibility is based on his particular approach to causation 'according to which our interventions in the world are cut short by later interventions, so that each person has a set of achievements or failures that is exclusively their own or is shared with a few others'.⁹⁴ He considers that '[h]ow far one should go back in a causal inquiry depends on the purpose of the inquiry'.⁹⁵ This introduces an element of circularity to Honoré's account as Cane highlights:

For the purposes of attributing responsibility, then, principles of causation are inextricably linked with principles of responsibility. And just as what counts as a cause depends on the purpose of the causal inquiry, so the answer to the question of what we are responsible for will depend on the purpose of the inquiry...In short, the scope of outcome responsibility...depends on the purposes it serves.⁹⁶

It is clearly not possible for outcome responsibility to determine objectively which outcomes a person is responsible for. Instead, outcome responsibility requires a value-judgment about which outcomes a person ought to be considered responsible for, and it is this normative aspect that needs to be justified. So to say that the doctor in *Chester* was morally responsible for the harm caused by his carefully executed operation on the basis of outcome responsibility is merely to say that his action was a cause of the harm and the purpose of the vindicating the patient's right to know/enforcing the doctor's duty to warn through the negligence inquiry justifies the attribution of responsibility. Effectively this amounts to saying that *if* the purpose of the inquiry is vindication/deterrence rather than corrective justice then there is a reason for concluding that

⁹³ Peter Cane, 'Responsibility and Fault: A Relational and Functional Approach to Responsibility' in Peter Cane and John Gardner (eds.) *Relating to Responsibility, Essays in Honour of Tony Honoré on his 80th Birthday* (Hart Publishing 2001) 91.

⁹⁴ Tony Honoré, *Responsibility and Fault* (Hart Publishing 1999) 10.

⁹⁵ *ibid* 11.

⁹⁶ Cane, 'Responsibility and Fault' (n93) 90.

the doctor is morally responsible for the outcome. It is not the force of his moral responsibility for the outcome that justifies the imposition of liability.⁹⁷

In corrective justice a person is not responsible for all loss that they cause, only for wrongful loss. Causation thus provides the essential link between the parties so it is the starting point for being able to find moral responsibility but its role is limited because it cannot explain where moral responsibility for outcomes ends. It is the concept of wrongdoing that provides the dividing line between those losses that are wrongful and those that are not, and this must be addressed through the remaining negligence doctrines other than causation.

2. The doctrinal framework

Having established that negligence should be seen as a corrective justice-based system, i.e. a system premised on interpersonal responsibility and a conception of morality that takes seriously the intrinsic moral worth of people, and having seen the place of causation within interpersonal responsibility, it is important to turn to the question of how this is translated into practice through the doctrines of negligence law. As explained in the introduction to this chapter, many problems that arise in exceptional approaches to causation exist because the doctrine of causation is used to address other aspects of responsibility which are not truly causal. It is essential to clarify ‘what causation is not’ and to illustrate clearly that evaluative aspects of responsibility are addressed through the other negligence doctrines.

Coherence ‘signifies a mode of intelligibility that is internal to the relationship between the parts of an integrated whole’,⁹⁸ such that ‘[t]he negligence concepts form an ensemble that brackets and articulates a single normative sequence’.⁹⁹ Having shown that causation has a vital yet limited role within the concept of interpersonal responsibility it is essential that this is reflected in the negligence doctrines. The overall objective of this section is therefore to show the

⁹⁷ The relationship between causation and responsibility will be revisited in chapter two on causation.

⁹⁸ Weinrib, *The Idea of Private Law* (n59) 14.

⁹⁹ Weinrib, ‘Legal Formalism’ (n55) 593.

interrelationship between causation and the other negligence doctrines in order to emphasise the role played by the other doctrines in determining liability together.

2.1 The duty of care

The duty of care dominates this section because it is there that the issues of interpersonal responsibility and wider distributive arguments come to the fore. It was argued in section 1.4 that it is theoretically sound to maintain that liability should only be imposed if it is grounded in considerations of interpersonal responsibility although it may exceptionally be limited to avoid distributive injustice. This section therefore proceeds to flesh out this argument by evaluating the concrete issues that motivate duty decisions.

2.1.1 Duty as the ‘central organising concept’¹⁰⁰

Coherence requires that the negligence doctrines come together to form an integrated whole. A virtue of Beever’s corrective justice-based approach to negligence is that it seeks to achieve this:

[T]he ‘principled approach’...treats the stages of the negligence enquiry as forming a unified investigation. This investigation is an enquiry into whether the unreasonable risk created by the defendant (standard of care) was an unreasonable risk of injury to the claimant (the duty of care) and an unreasonable risk of the injury that the claimant suffered (remoteness).¹⁰¹

Treating the negligence doctrines as forming a ‘unified investigation’ reflects Weinrib’s concern, noted above, that the negligence doctrines ‘articulate a single normative sequence’.¹⁰² However, Beever addresses the standard of care before he addresses the duty of care. The danger of this is that the duty doctrine risks appearing as a control mechanism rather than as a meaningful concept and this does not sit comfortably with a moral account of negligence liability.

¹⁰⁰ This label is used by Nicholas J McBride, ‘Duties of care – do they really exist?’ (2004) 24 OJLS 417, 423-24.

¹⁰¹ Beever, *Rediscovering* (n10) 115.

¹⁰² Weinrib, ‘Legal Formalism’ (n55) 593.

If the duty doctrine is merely a control mechanism, or is merely a duty to compensate a victim if one's carelessness results in loss, it would lose its moral force. McBride explains that in this kind of 'cynical' view of duty, '[t]he law of negligence would no longer purport to guide people's behaviour...Instead, the law of negligence would simply attach costs to certain forms of behaviour and leave it up to individual citizens to decide whether or not to participate in those forms of behaviour'.¹⁰³ This cynical view is incompatible with Kantian Right since it treats the claimant (and potential claimants) merely as a commodity or as a cost to be entered into an equation. Corrective justice is a moral theory so it implies an idealist approach to duty as a meaningful duty to take care since this takes seriously the moral worth of people.

McBride goes on to argue that the duty is the 'central organizing concept around which the whole of the law of negligence revolves' in an idealist account:

After all, an idealist would say, a claimant will only be entitled to sue a defendant in negligence if: (1) the defendant owed her a *duty of care*; (2) the defendant breached that *duty of care*; (3) the breach of that *duty of care* caused her to suffer some kind of loss; and (4) the loss suffered by the claimant as a result of the defendant's breach was the kind of loss which the *duty of care* breached by the defendant was imposed on him in order to avoid.¹⁰⁴

Logically, therefore, duty must be analysed before the other doctrines. This is important because the arguments raised at the duty stage focus on the relationship between the parties, so they bring to the foreground the considerations of interpersonal responsibility in that particular relationship. Where problems arise in the remaining doctrines, and a solution is sought that forms part of a coherent approach to negligence, then we must go back to the issues of interpersonal responsibility identified at the duty stage and carry them through into the solution to the problem.

McBride illustrates the centrality of the duty of care in establishing the interpersonal aspect of responsibility and its role in informing the remaining negligence doctrines through cases such as

¹⁰³ McBride (n100) 424.

¹⁰⁴ McBride (n100) 423-24.

Reeves v Commissioner of Police of the Metropolis.¹⁰⁵ This case involved the duty of care owed by the police to a prisoner in their custody to take reasonable care to prevent him committing suicide. The police failed to take reasonable care and the prisoner took his own life. Ordinarily the voluntary and unreasonable act of the victim in harming himself would be considered at the legal causation stage of the negligence inquiry to have broken the chain of causation between the defendant's negligence and the victim's harm. But in the context of this particular duty of care, a duty specifically targeted towards preventing the victim from self-harm, the victim's voluntary and unreasonable act of self-harm cannot be considered to break the chain of causation. This is not merely a matter of reinforcing the duty owed by the police in the way that imposing liability on the doctor in *Chester v Afshar* in the absence of a causal link was directed towards enforcing the content of the doctor's duty to warn. In *Reeves* there *was* a causal link, and the issues of interpersonal responsibility showed that the coherent solution was to hold the harm to fall within the scope of this responsibility. The duty of care similarly informs the remaining negligence doctrines in cases where harm is inflicted by a third party.¹⁰⁶

Davies also draws on the idea that the duty of care and the remaining negligence doctrines form unified investigation in order to explain the evolution of the test for remoteness.¹⁰⁷ He suggests that the approach adopted in *Re Polemis and Furness Withy Co*,¹⁰⁸ asking whether the harm was a direct consequence of the negligence regardless of whether it was reasonably foreseeable, was acceptable given the narrow approach courts took to duty of care at that time. Since *Re Polemis* was decided before *Donoghue v Stevenson*,¹⁰⁹ the duty relationship had to be shown to fall within an established duty situation, what Davies refers to as a 'particularised, relational' duty of care,¹¹⁰ so

¹⁰⁵ *Reeves v Commissioner of Police of the Metropolis* [2000] 1 AC 360 (HL).

¹⁰⁶ Claire McIvor, *Third Party Liability in Tort* (Hart Publishing 2006) 18-19.

¹⁰⁷ Martin Davies, 'The Road from Morocco: Polemis through Donoghue to no-fault' (1982) 45 MLR 534.

¹⁰⁸ [1921] 3 KB 560.

¹⁰⁹ [1932] AC 562 (HL(Sc)).

¹¹⁰ *ibid* 541. In this instance, Davies says, the duty was based on the authority of *Dean v Keate* (1811) 3 Camp. 4 (170 ER 1286) which established that a bailee owes the bailor of property a duty to exercise reasonable care over the property he has hired.

‘a wide remoteness test was unexceptionable when the duty was narrowly conceived’.¹¹¹ This remoteness test of ‘directness’ was later overruled in *The Wagon Mound*¹¹² in favour of a test of reasonable foreseeability of the type of harm. This shift, Davies suggests is explained in part by the expansion of the duty of care in *Donoghue v Stevenson*. The more restrictive nature of the *Wagon Mound* test was better suited to limiting liability given the less restrictive stance to establishing duty of care which was then being taken applying Lord Atkin’s ‘neighbour principle’. Given the more restrictive incremental approach that is now being taken towards duty of care under the *Caparo* test,¹¹³ and the return to what Stapleton labels ‘pockets of liability’,¹¹⁴ this may explain why courts are beginning to look towards the ‘scope of duty’ as the approach to remoteness.¹¹⁵ Such a view is supported by Steele who has suggested ‘[i]n more recent years, the concept of ‘duty’ has become more specific to the particular circumstances. Perhaps it is time to update *The Wagon Mound*, to reflect this change in the approach to duty’.¹¹⁶

Having established that the duty of care is central to shaping the rest of the negligence inquiry because of the role it plays in identifying the particulars of the relationship of interpersonal responsibility between the parties to the case, it is important to turn to the duty itself and consider which duty factors are truly reflective of interpersonal responsibility.

2.1.2 Identifying issues of interpersonal responsibility in the duty doctrine

This section examines the duty doctrine to identify those duty factors that reflect interpersonal responsibility and ought to influence the approach to liability questions including causation/damage, and to filter out those that reflect social welfare/distributive justice and are therefore incapable of grounding liability. It was argued in the theory section that this approach

¹¹¹ Davies (n107) 541. He also argues that the social context of *Re Polemis* made the decision desirable at the time.

¹¹² *Overseas Tankship (UK) Ltd v Morts Dock and Engineering Co Ltd (The Wagon Mound (No1))* [1961] AC 388 (PC)

¹¹³ *Caparo Industries v Dickman* [1990] 2 AC 605 (HL).

¹¹⁴ Jane Stapleton, ‘Duty of care and economic loss: a wider agenda’ (1991) 107 LQR 249.

¹¹⁵ *South Australia Asset Management Corp (SAAMCo) v York Montague Ltd* [1997] AC 191 (HL), 210-13 (Lord Hoffmann).

¹¹⁶ Jenny Steele, *Tort Law: Texts, Cases and Materials* (2nd ed., Oxford University Press 2011) 211.

is appropriate, so the current section seeks to flesh out the various considerations to show what it means in concrete terms. This is because it is the issues of interpersonal responsibility that will be prioritised when considering exceptional approaches.

The dividing line between considerations of interpersonal responsibility and distributive factors is sometimes said to be the line between ‘principle’ and ‘policy’, or the distinction between ‘foreseeability and proximity’ and ‘fairness, justice and reasonableness’. Since these terms do not have any agreed meaning it is necessary to give concrete examples of a range of considerations that do/do not reflect interpersonal responsibility.

Beever distinguishes between ‘principle’ and ‘policy’ although this division has been criticised by legal realists, notably Stapleton, who argue that there is no distinction to be drawn and that reasons for liability will be labelled ‘principle’ to make them seem palatable and ‘policy’ to make them seem political. Stapleton argues that ‘appeal merely to so-called ‘principles’ often simply masks the substance of a judge’s reasoning process’.¹¹⁷ Witting has similarly criticised Beever’s corrective justice-based account of negligence on the basis that ‘Beever provides very little comfort to those who prefer principle when ‘principle’ is little more than Beever says it is – namely that law as determined by Beever’s own application of reflective equilibrium’.¹¹⁸ However Beever’s exposition of ‘principle’ is not nearly as subjective as Witting suggests. Beever explains that:

‘Principle’ refers to the rules and doctrines of the law itself... ‘Policy’, on the other hand, can be defined only negatively. ‘Policy’ is everything apart from principle. For example, policy has been held to include issues of distributive justice, social morality, economic efficiency, public opinion and so on. But it is impossible to define the content of these terms exactly, because people disagree on what constitute the rules and doctrines of the

¹¹⁷ Jane Stapleton, ‘The Golden Thread as the Heart of Tort Law: Protection of the Vulnerable’ (2003) 24 Aust Bar Rev 135, 136.

¹¹⁸ Christian Witting, ‘The House that Dr Beever Built: Corrective Justice, Principle and the Law of Negligence’ (2008) 71 MLR 621, 630.

law. Accordingly, there is no ‘theory neutral’ way of defining the content of principle and policy.¹¹⁹

In other words, the distinction between principle and policy is dependent on the foundational theory of negligence that one prefers. So ‘principle’ is not simply ‘what Beever says it is’; ‘principle’ is the set of rules that express corrective justice. To an extent this is subjective because it depends on Beever’s account of what corrective justice is. But it does not suffer from the indeterminacy and subjectivity that Witting implied. Indeed, although Witting does not necessarily draw such a clear line between principle and policy, he has argued elsewhere that there is a meaningful distinction to be drawn between the three elements of the *Caparo* test for duty of care, notably between ‘proximity’ and consideration of what is ‘fair, just, and reasonable’. He suggests that proximity looks inwards to the relationship between the parties and ‘fair, just and reasonable’ looks outwards to ‘the potential impact that the recognition of a duty might have upon a number of parties who are *not before the court*’.¹²⁰ This is broadly correct but the dividing line is not quite so clear and bright.

Robertson’s account is particularly useful, especially since he advocates the broad approach adopted in this thesis. Robertson prioritises justice between the parties but, in contrast to Beever whose pure corrective justice account is ‘designed to eliminate both the need for appeal to policy considerations and the permissibility of that appeal’,¹²¹ Robertson ‘recognises that community welfare considerations can come into play in a secondary way to negate the duty of care in certain limited circumstances’.¹²² He draws out considerations across all three stages of the *Caparo* test that address ‘justice between the parties’ and considerations at each stage that concern ‘community welfare’.¹²³ In other words, he draws a broad distinction between ‘justice between

¹¹⁹ Beever, *Rediscovering* (n10) 3.

¹²⁰ Christian Witting, ‘Duty of care: an analytical approach’ (2005) 25 OJLS 33, 38.

¹²¹ Beever, *Rediscovering* (n10) 29.

¹²² Roberston, ‘Rights, Pluralism and the Duty of Care’ (n5) 441.

¹²³ *ibid* 441-4.

the parties' which is the focus of corrective justice-based interpersonal responsibility, and 'community welfare' which is distributive.

In terms of justice between the parties, Robertson says that there are three 'stages': reasonable foreseeability, proximity, and 'other considerations of justice between the parties'. Reasonable foreseeability is widely regarded as a necessary element since it signals that the defendant at least had capacity to act differently – it makes the first step away from purely causal responsibility towards a form of moral responsibility for outcomes. 'Proximity' is not susceptible to objective definition, but as Robertson argues 'even if the concept of proximity does no more than to give focus to a particular aspect of the duty inquiry and to express the nature of what is in issue in the assessment of a discrete and distinctive set of considerations, that may be considered enough to justify its retention'.¹²⁴ Using similar terms to Lord Atkin's formulation in *Donoghue v Stevenson*, he explains that proximity 'is essentially concerned with the question whether the claimant was so closely and directly affected by the defendant's conduct that the defendant ought to have had the claimant in mind when considering whether and how to act'.¹²⁵ Proximity factors include:

the extent of the defendant's control over the risk, whether the defendant created the risk, the extent of the defendant's knowledge of the risk, the magnitude of the risk and degree of foreseeability of the harm, the extent of the claimant's vulnerability to harm arising from the risk, the extent to which the claimant relied on the defendant, whether the defendant knew of that reliance and whether the defendant can be said to have assumed a responsibility, either to the claimant or for a particular task.¹²⁶

Robertson also notes that proximity allows the court to look at the scope and limits of the claimant's personal responsibility and the defendant's autonomy. The availability of insurance is irrelevant here because as Robertson explains it is an alternative compensatory remedy which does not actually render the claimant less vulnerable to harm, just more able to deal with the

¹²⁴ Andrew Robertson, 'Justice, community welfare and the duty of care' (2011) 127 LQR 370, 374.

¹²⁵ ibid 375.

¹²⁶ ibid.

financial consequences of harm. The ‘other considerations of justice between the parties’ that Robertson identifies are first whether a duty in negligence would be inconsistent with the contractual context between the parties, and second whether recognition of a duty would impose an unreasonable burden on the defendant. This second consideration seems most controversial, because as Stapleton has suggested, the fact that a defendant might be exposed to a large (as opposed to indeterminate) number of claims is an unconvincing reason to deny liability.¹²⁷ The unreasonableness of the burden imposed on the defendant in Robertson’s view is less a function of the number of claims/extent of loss inflicted, and more a function of the steps required for the defendant to fulfil his duty since Robertson focuses on liability for omissions or for acts of third parties. There seems to be a difference between what Robertson would think of as a burden and what Stapleton would think of as a burden which probably arises from their different views of what the duty concept is for – since Stapleton sees it as a policy-based doctrine she treats it as a control device, whereas Robertson treats it as an actual duty.

The community welfare concerns that Robertson then raises, which have a secondary importance, are justiciability, community welfare, and other ‘non-justice’ factors. Justiciability concerns judicial competence, judicial deference to legislature and executive, and cases where the court would not be able to set a standard of care e.g. a standard owed by a pregnant woman to her unborn child. Robertson suggests that where an issue is not justiciable this will always trump considerations of justice between the parties. Community welfare factors include any conflict between the duty to be owed to the claimant and duties the defendant already owes to others, inconsistency within the legal system, availability of more effective/efficient remedies elsewhere, effect on the community e.g. uninsurability, strain on the public purse, flood of unmeritorious claims, inefficient allocation of risk, over-deterrence/defensive practice. Finally the ‘other non-justice factors’ to which he refers include deontological considerations such as refusing a duty owed by participants to one another in a joint criminal enterprise because recognition of a duty

¹²⁷ Jane Stapleton, ‘Duty of Care Factors: a Selection from the Judicial Menus’ in Peter Cane and Jane Stapleton (eds.) *The Law of Obligations: Essays in Celebration of John Fleming* (Clarendon Press 1998) 66.

would condone breaches of criminal law. Community welfare considerations alone cannot justify liability which must be based on the demands of justice between the parties, but if there are convincing community welfare arguments against liability then pro-duty community welfare factors should also be weighed up.

The problem is that the introduction of these community welfare arguments introduces an element of indeterminacy because, as we have seen, community welfare arguments introduce an element of pluralism where no single factor has priority over the others.¹²⁸ As argued earlier, however, coherence should not be pursued to the detriment of morality, so this small element of indeterminacy is a necessary complication. The difficulty then is in drawing a line and ensuring that community welfare factors are only exceptionally allowed to impact on liability. Beever has objected to policy arguments, arguing that '[o]ver the years, [policy] has grown from a small suburb outlying the town of legal principles into a metropolis that now dwarfs and encroaches upon the town.¹²⁹

The ubiquity of policy, or community welfare considerations, is problematic because of the indeterminacy inherent in it. Beever seeks to resolve this indeterminacy by adhering to a strict corrective justice view of negligence: 'viewing the law of negligence in terms of the principled approach is designed to eliminate both the need for appeal to policy considerations and the permissibility of that appeal'.¹³⁰

The approach advocated in this thesis allows for consideration of community welfare concerns but only exceptionally whilst prioritising matters of justice between the parties. To adopt Beever's analogy, it seeks to restore corrective justice to a metropolis and confine distributive justice to a small suburb.

The final part of this chapter will outline the relationship between causation and the remaining negligence doctrines. There is a danger that where causation is a problem it is thought of as

¹²⁸ See for example Robin Oppenheim, 'The "mosaic" of tort law: the duty of care question' [2003] J P I Law 151, 165.

¹²⁹ Beever, *Rediscovering* (n10) 3.

¹³⁰ *ibid* 29.

determinative of liability. As well as considering the duty issues in the case courts must remember that the other doctrines play their part in defining the contours of liability. It is therefore important to know what these doctrines consist of and how they are related to causation.

2.2 Actionable damage/the gist of the negligence action

A claimant must have suffered an actionable form of damage in order to bring a successful negligence action; damage is said to form the gist of the action. But the question of what constitutes actionable damage is often overlooked for the simple reason that in most cases a claimant will have suffered clearly observable physical damage to his person or property. The definition of damage is inextricably linked to the causation inquiry as Stapleton insists:

It cannot be over-emphasised that the formulation of the ‘damage’ forming the gist of the action *defines* the causation question. Logically one can only deal with causation after one knows what the damage forming the gist of the action is.¹³¹

Both damage and causation are then related to the later question of quantification, or valuation, of the claimant’s loss. One theme that will recur in this thesis is that the three concepts of damage, causation, and quantification though related must remain analytically distinct. This is especially important where claimants have attempted to overcome difficulties of proof of causation by reconceptualising the damage as the loss of a chance of a better outcome, or an increase in the risk of harm.¹³² This section prepares the groundwork by briefly explaining the concept of damage before outlining its relationship to causation and quantification so that the basic issues are understood in advance of the more detailed analysis in later chapters.¹³³

¹³¹ Jane Stapleton, ‘The Gist of Negligence: Part 2 The Relationship Between ‘Damage’ and Causation’ (1988) 104 LQR 389, 393.

¹³² E.g. *Hotson v East Berkshire Health Authority* [1987] AC 750, 3 WLR 232 (HL); *Gregg v Scott* [2005] UKHL 2, 2 AC 176; *Barker v Corus* [2006] UKHL 20, 2 WLR 1027. These attempts to reformulate the damage form the focus of Chapters Three and Four.

¹³³ Chapter Three addresses loss of a chance and Chapter Four, which addresses the evidentiary gap, considers a range of responses including reformulating actionable damage as the risk of harm rather than the physical outcome.

2.2.1 Damage

As explained, corrective justice concerns wrongful loss in interactions. The question of whether a loss is wrongful depends partly on whether it was caused by wrongdoing, and also on whether it is legally recognised as constituting actionable damage. Property damage, some personal injury, and financial loss are recognised forms of damage (although with regard to purely financial loss, the circumstances in which one is owed a duty of care are much more limited). Conditions such as anxiety and grief that fall short of medically recognised psychiatric illness are not actionable.¹³⁴ Similarly, not all physical illness will constitute actionable damage: it was held in *Rothwell* that although asymptomatic pleural plaques are a medically recognised condition they do not constitute actionable damage.¹³⁵ Lord Hoffmann explained that damage ‘is an abstract concept of being worse off, physically or economically, so that compensation is an appropriate remedy. It does not mean simply a change in physical condition’.¹³⁶

It is important to identify the actionable damage clearly at the outset of any claim so that the causation inquiry can be directed to this damage. One issue that will be addressed in greater detail in Chapter Two is the distinction between ‘divisible’ and ‘indivisible’ damage. These terms are not used consistently amongst academics. The appropriate definition will be discussed in detail in that chapter, but broadly speaking damage is said to be divisible if it is dose-related whereas indivisible damage is ‘all or nothing’.

2.2.2 Quantification

At the stage of quantification of the claimant’s loss, where the task is to return him to his pre-tort position, causal concepts are invoked but the question is different. At the factual causation stage of the negligence inquiry, the question is whether the defendant’s negligence was a cause of

¹³⁴ *Hinz v Berry* [1970] 2 QB 40 (CA); *Rothwell v Chemical and Insulating Co Ltd* [2007] UKHL 39, [2008] 1 AC 281.

¹³⁵ *Rothwell* (n134).

¹³⁶ *ibid* [7] (Lord Hoffmann).

the damage suffered by the claimant. At the quantification stage, the court must ask whether the claimant is any worse off because of this physical damage.

In *Performance Cars v Abraham*,¹³⁷ the claimant's car had been damaged in a collision due to the defendant's negligence. Among the damage caused to the car was damage to the front wing which, due to the nature of the paint, necessitated the re-spraying of the whole lower part of the car. However, in a previous collision the rear wing had been damaged which also necessitated the re-spraying of the whole lower part of the vehicle. Although the claimant had obtained judgment against the party responsible for the first accident, the judgment had not been satisfied and he sought to recover the cost from the defendant responsible for the second collision. The Court of Appeal held that although the defendant had caused physical damage to the front wing of the car, the cost of re-spraying the lower part could not be recovered from the defendant because the pre-tort state of the car was that the lower part already needed re-spraying. In other words, the defendant was a cause of physical damage to the car but at the valuation stage it was found that the physical damage had not caused the claimant to be any worse off. This is not a question of causation – it is clear that the defendant's negligence was a factual cause of the physical damage – but with respect to this particular portion of the loss associated with the physical damage the value is zero because the car already needed re-spraying before the physical damage occurred.

At the quantification stage courts must also determine what course the claimant's life would otherwise have taken and whether other factors would have led to him incurring similar loss in the future anyway. This can be illustrated through *Smith v Leech Brain*.¹³⁸ In this case the defendant's negligence caused the claimant to suffer a burn on his lip which triggered a pre-cancerous condition and led to his eventual death. As a matter of factual causation there was no doubt that the defendant's negligence had caused the claimant to suffer the harm at that particular time. But in quantifying the value of the loss caused by the defendant, the court made

¹³⁷ *Performance Cars v Abraham* [1962] 1 QB 33 (CA).

¹³⁸ *Smith v Leech Brain & Co* [1962] 2 QB 405 (QBD).

a reduction in damages to account for the fact that the claimant had a pre-cancerous condition so was likely to have died prematurely anyway. Under the ‘vicissitudes of life’ principle, courts ordinarily must reduce the value of damages to take account of the likelihood that the claimant would have suffered similar loss at some stage in his life, and where there is information specific to the claimant allowing them to personalise this calculation then they will do so. This ‘reduction’ in damages does not reflect any doubt that the defendant’s negligence was a cause of the claimant’s loss, but reflects the attempt to value that loss as accurately as possible by calculating how likely it was that the claimant would have suffered similar misfortune at some stage in his life because of independent events.

The question of what course the claimant’s life would have taken, and how far the defendant’s responsibility extends, is informed by factual causation but it is also a qualitative inquiry that involves evaluating the extent of the loss that is ‘wrongful’. This is apparent in the contrast between the solutions in *Jobling*¹³⁹ and *Baker*.¹⁴⁰ In terms of the causal processes, *Jobling* and the American case of *Dillon*,¹⁴¹ seem closer to *Smith v Leech Brain* than *Baker* does. In *Jobling* the defendant caused the claimant a back injury which reduced his earning capacity but before reaching the court the claimant suffered a naturally occurring back illness which would have prevented him from working anyway – although the negligence did not trigger this underlying condition as had been the case in *Smith* the court was still able to accurately calculate the vicissitudes of life for this claimant by the time of trial. At the time of the negligence there was already some idiosyncrasy of the claimant that meant that he was destined to suffer this loss. Likewise in *Dillon*, as the claimant was falling to his death he was electrocuted due to the defendant’s negligence – at the time of the negligence he already had a (significantly) reduced life expectancy. In contrast, in *Baker* the claimant’s leg was damaged by the defendant’s negligence but at that time there was no train of events already in motion that meant that he was destined to

¹³⁹ *Jobling v Associated Dairies Ltd* [1982] AC 794 (HL).

¹⁴⁰ *Baker v Willoughby* [1970] AC 467 (HL).

¹⁴¹ *Dillon v Twin State Gas & Electric Co*, 85 NH 449, 169 A 111 (1932)

suffer leg injury in the future. Yet by the date of trial he had been shot in that leg by robbers. This is different from *Jobling* and *Dillon* in the sense that there was no risk of being shot particular to the claimant at the time of the defendant's negligence, so restoring him to his pre-tort position is to return him to a position where he should have continued to have full use of his leg (in *Jobling/Dillon* at the time of the negligence there was *already* a reason why the claimant would not continue to have full use of the relevant body part in future). But this issue of quantification also reflects the court's qualitative assessment of the extent to which the claimant's loss was wrongful, with the concern in *Baker* that the claimant had been exposed to two careless acts and his claim should not be allowed to 'fall between two stools'.

2.3 Breach of duty

Just as the definition of damage is inextricably linked to the causation question, so is the framing of the breach of duty since it must be established that the defendant's negligence in particular, rather than the defendant's conduct in general, was a cause of the damage suffered. Stapleton has illustrated this through her analysis of the decision in *McWilliams v Arrol*.¹⁴² The claimant fell to his death from scaffolding and the defendant employer was negligent in failing to provide a safety harness. To prove that failure to provide a harness was a cause of the death, it is necessary to prove that the victim would have worn a harness if it had been provided, and the claimant was unable to prove this. If, however, reasonable care required not only the provision of safety harnesses but also instating a surveillance system to ensure that safety equipment was used, then this breach of duty would be a cause of his death.¹⁴³

The breach and causation inquiries are further interlinked in cases such as *Bolitho* where the answer to the question of whether the damage would have occurred if the defendant had not breached his duty depends on an assessment of the hypothetical course of action that the

¹⁴² *McWilliams v Sir William Arrol Co Ltd* [1962] 1 WLR 295 (HL).

¹⁴³ Jane Stapleton, 'Cause in Fact and the Scope of Liability for Consequences' (2003) 119 LQR 388, 391-2.

defendant would then have followed.¹⁴⁴ The victim in *Bolitho* was a child who suffered brain damage and later died as a result of respiratory failure. He experienced two episodes of respiratory difficulties and the doctor was negligent in failing to attend the patient when called. The defendant argued, however, that respiratory failure would only have been avoided by intubating the patient and if she had attended she would not have intubated, so her failure to attend did not cause the injury. The House of Lords confirmed that courts should go on to ask whether this hypothetical conduct would also have been negligent; a defendant cannot escape liability by proving that his negligence made no difference to the outcome because he would still have caused the harm by a later act of negligence.¹⁴⁵ This reflects the notion that the negligence doctrines ‘articulate a single normative sequence’.¹⁴⁶

This chapter has shown that it is important not only for the doctrine of causation to be internally intelligible but also for it to form part of an overarching approach to the tort of negligence that is also coherent. Aristotelian corrective justice and Kantian Right, it was argued, provide a coherent theoretical foundation that is not only formal but also implies a moral account of wrongdoing that expresses interpersonal responsibility. The various negligence doctrines should come together to give effect to this corrective justice-based interpersonal responsibility. The duty of care is central to the negligence inquiry; it is at the duty stage that the courts will identify the issues of interpersonal responsibility that should shape application of the remaining doctrines so that they ‘articulate a single normative sequence’. Causation is an essential element of interpersonal responsibility because it is the causal relationship between the wrongdoing and the loss that joins the defendant and claimant in an interaction that has resulted in a ‘wrongful loss’.

¹⁴⁴ *Bolitho v City of Hackney Health Authority* [1998] AC 232 (HL).

¹⁴⁵ The Court of Appeal later clarified in *Gouldsmith v Mid-Staffordshire General Hospitals Trust* [2007] EWCA Civ 397, [2007] LS Law Medical 363, that there are two stages to this inquiry: first, what hypothetical course of action would the defendant/third party probably have taken?; second, if they would not have given harm-preventing treatment would this have been negligent? If the answer to the first question is that the defendant/third party probably would have undertaken treatment to prevent the harmful outcome then it is clear that their initial negligence did cause the harm and it is not necessary to prove that they would have been bound, in the *Bolam/Bolitho* sense, to follow this hypothetical course of action.

¹⁴⁶ Weinrib, ‘Legal Formalism’ (n55) 593.

However, causation also has a limited role and the evaluative elements of responsibility must be addressed through the remaining negligence doctrines. These doctrines also help frame the causation question because they identify the loss and the wrongdoing between which the causal relationship is sought.

This foundation means that the analysis of causation that will be developed in the remaining chapters is based on a clear understanding of the underlying theory of negligence law and of the demands that negligence makes of the doctrine of causation.

Chapter Two: Identifying the Proper Function of Causation

Introduction

As established in chapter one, causation is central to interpersonal responsibility, but the law relating to causation is currently in a state of confusion. This confusion extends not only to the exceptional approaches to causation, but characterises the standard approaches as well. This chapter will attempt to clear away the confusion by adopting a clear approach to understanding the causal problems as well as their solutions. The starting point to developing any test for causation must be an accurate understanding of what purpose the doctrine of causation has in negligence. It will be argued here that the concept of causation in law is narrower and more precise than the use made of the term ‘cause’ in everyday language because, as Stapleton has explained, causation can be understood in a wide range of senses, from involvement, to explanation, to blame.¹ Moreover, as Wright, building on the work of Hart and Honoré, argues, philosophy provides the law with a robust account of causation that ought to form the basis of any legal test.² By understanding the philosophical account of causation developed by Hume and Mill, we can visualise and better understand the causal problems that come before the courts. This will enable the law to distinguish between those cases that are straightforward, and those that are actually problematic. Negligence law does, however, put its understanding of causation to a specific purpose, namely establishing a causal link between a particular instance of negligence and a particular harmful outcome.³ A certain level of pragmatism must therefore be allowed in order to avoid becoming distracted by details that affect philosophers and scientists. Effectively, we all use a shared concept of causation, derived from philosophy, but we must retain a clear vision of what demands the law places on a test of causation.

¹ Jane Stapleton, ‘Choosing what we mean by “Causation” in the Law’ (2008) 73 Mo L Rev 433, 438.

² Richard Wright, ‘Causation in Tort Law’ (1985) 73 Cal L Rev 1735 (hereafter ‘Causation’), 1774-1812.

³ Richard Wright, ‘The NESS Account of Natural Causation: A Response to Criticisms’ in Richard Goldberg (ed.) *Perspectives on Causation* (Hart Publishing 2011) 290-1.

The philosophical account of causation is best reflected by Richard Wright's NESS test rather than the but-for test that currently dominates.⁴ The but-for test is mismatched with the philosophical reality of what causation is and this means that the use of the but-for test has created apparent problems that do not, in reality, exist. The legal solutions to these problems, particularly the 'material contribution to harm' test, have created more difficulties than they have solved. The reasons for this are twofold. First, because causation is approached from the paradigmatic mind-set of the but-for test which is not an accurate reflection of the meaning of causation, we are unable to conceptualise the problems that exist. This has contributed to the second difficulty regarding the emergence of a vocabulary of causation-specific terminology: divisibility of disease, cumulative causes, alternative causes. These terms are technical so they ought to have a technical meaning but the difficulties of conceptualising the causal problems to which they relate means that they do not have any agreed meaning. They are used to mean different things by different people, thus exacerbating the confusion. By beginning again from the most basic causal problems and explaining them through the conceptual framework of the NESS test, many of these problems can be avoided.

This chapter will therefore begin by asking how the legal concept of causation relates to causation in the fields of everyday language, and philosophy and science. After explaining the philosophical account of causation, it addresses the question of how best to translate this into a legal test. Both the but-for test and the NESS test will be applied to a range of theoretical problem scenarios to illustrate how the NESS test is able to solve problems that the but-for test cannot. Finally, the NESS test will be applied to existing case law to show how it allows us to reconceptualise the problems and to understand the specific terminology.

⁴ Wright, 'Causation' (n2); Richard Wright, 'Causation, Responsibility, Risk, Probability, Naked Statistics, and Proof: Pruning the Bramble Bush by Clarifying the Concepts' (1987) 73 Iowa L Rev 1001 (hereafter 'Pruning'); Wright, 'The NESS Account' (n3). Wright builds on the idea of a 'causally relevant factor' first developed by Hart and Honoré, *Causation in the Law* (Clarendon Press 1959).

1. The demands on causation

Before it is possible to say what qualities are required from a test for causation, we must understand what is actually meant by ‘causation’. More importantly, we need to know what we, as lawyers, mean by causation. What role does it fulfil in the negligence inquiry, and does it have a special ‘legal’ definition? In the previous chapter it was shown that causation has a pivotal yet limited place in corrective justice, and the two strands of the current section will unpack what this implies for the doctrine of causation within the negligence inquiry. First, causation has a limited role in attributions of responsibility so as Wright argues, in comparison with the use of the word ‘causation’ in everyday language, there is a technical legal meaning.⁵ ‘Causation’ in everyday language is used flexibly to mean different things in different circumstances. In contrast, the law requires greater consistency and precision in the definition of causation. The factual causation element of negligence is concerned with causation in a purely factual sense, what Hart and Honoré call being a ‘causally relevant condition’, or Stapleton calls ‘involvement’. The second part of this section addresses the point that causation is essential to interpersonal responsibility meaning that the doctrine of causation must be conceptually robust. As argued first by Hart and Honoré, and subsequently by Wright, the most complete account of factual causation can be found in the philosophy of Hume and Mill, and this should inform the legal doctrine of factual causation.

1.1 The limited role of causation in negligence

It has been suggested that the doctrine of causation should reflect the ordinary man’s use of the word, for example in *McGhee v National Coal Board*, Lord Reid said ‘it has often been said that the legal concept of causation is not based on logic or philosophy. It is based in the practical way in

⁵ Indeed Wright suggests that although Hart and Honoré’s philosophical account of causation was ‘a major advance in the analysis of causation in both law and philosophy’ it was ‘overshadowed and distorted by their primary emphasis on elaborating supposedly factual ‘common sense’ principles for treating only some causally relevant factors as causes, so that, initially, it received minimal attention in the legal literature’. See Wright ‘The NESS Account’ (n3) 288.

which the ordinary man's mind works in the everyday affairs of life'.⁶ Indeed the seminal work of Hart and Honoré on causation sought to explain causation by reference to the way it is used in ordinary language.⁷ Wright has shown, however, that in everyday language 'cause' is often shorthand for 'the cause' rather than 'a cause':

The phrase "a cause" usually refers to causation per se – the fact of being one of many contributing conditions. The phrase "the cause" generally is used to denote which of the many contributing conditions is legally or morally responsible... "The cause" is merely an elliptical way of saying "the (most significant for our purposes) cause".⁸

In other words the ordinary man uses the word 'cause' to express a conclusion about responsibility or blame rather than limiting himself to observing a purely factual relationship of cause and effect. In comparison, the legal notion of factual causation is a technical concept since it refers to 'a cause' so care must be taken to separate the fact of causation from any attribution of responsibility that may be considered to flow from it.

A relatively straightforward example illustrates the distinction Wright draws.⁹ When a lit match is dropped and a fire results, the ordinary man might identify the dropping of the match as the cause. It is, however, only 'a cause', along with the presence of oxygen and flammable material. If a fire occurred during a particular manufacturing process that required the absence of oxygen, then common sense is to call the introduction of oxygen the cause of the fire. If a sofa does not meet safety standards concerning flammability then we might say that this, not the dropping of a match onto it, is the cause of the fire. In each of these scenarios 'the cause' is selected because it is the factor that is abnormal in the circumstances, and as Mackie has argued, '[w]hat is normal may depend upon man-made norms'.¹⁰ The identification of 'the cause' depends upon the purpose of the inquiry and the circumstances of the event. When the ordinary man determines

⁶ *McGhee v National Coal Board* [1972] 3 All E.R. 1008 (HL), 1011.

⁷ Hart and Honoré, *Causation in the Law* (2nd ed., Clarendon Press 1985).

⁸ Wright, 'Pruning' (n4) 1012.

⁹ This example is used by Hart and Honoré (n7) 11.

¹⁰ J. L. Mackie, *The Cement of the Universe: A Study of Causation* (Clarendon Press 1974) 119.

the cause of an event, such as a fire, he goes through two stages, the first factual and the second evaluative. The second of these stages is to pick out the particular factor that was unusual or blameworthy in the circumstances, which in the case of the fire above could be the match or the presence of oxygen or the flammability of the material depending on the circumstances. But first, although he did not necessarily consciously articulate it, he was aware that the conditions that must come together for a fire to start are oxygen, flammable material and a source of ignition. Each of these conditions is ‘a cause’ of a fire because they must each be present, but the ordinary man tends only to identify one as ‘the cause’, and which one he chooses will vary depending on the particular circumstances. As a test of factual causation it is therefore inappropriate for the law to rely on the ordinary man’s use of the word causation because the ordinary man makes an evaluative judgment without first articulating the factual judgment on which it is based.

The distinction between ‘a cause’ and ‘the (responsible) cause’ broadly corresponds to the division between factual and legal causation in negligence. Factual causation is truly causal, legal causation is an evaluative question of the appropriate extent of liability. Hart and Honoré, however, identified three issues and it is important to understand which of these are causal and which are evaluative. These three issues are: causally relevant condition, causal connection, and remoteness of damage.¹¹ The identification of ‘causally relevant conditions’ is the factual question of the historical involvement of the negligence in the occurrence of the harm.¹² This is what Wright referred to above as being ‘a cause’. The second question of ‘causal connection’ identifies which of the causally relevant conditions actually are, in Hart and Honoré’s terminology, causes and which are just conditions.¹³ It was in making the distinction between causes and conditions that Hart and Honoré turned to the use of the word ‘cause’ in everyday language. They argued that the ordinary man identifies something as a cause rather than a condition if it is non-coincidental, is free and deliberate human action or if it is something

¹¹ Hart and Honoré (n7) xlvii.

¹² ibid lvi, lxiv.

¹³ ibid 33.

abnormal. Finally the third issue considers whether the harm was too remote a consequence of the defendant's conduct. This, Hart and Honoré conceded, is a policy-based decision as to the appropriate scope of liability.¹⁴ In essence they identified the first stage as factual, the second as being based on principle and the third as based on policy.

For them, the first two issues are truly causal and the final issue is non-causal. This conclusion was probably driven by Hart and Honoré's aim of countering the Legal Realists' criticisms that causation is not neutral but is policy-driven.¹⁵ Hart and Honoré's work is important, among other reasons, for having shown that the task of identifying 'the cause' from among everything that could be called 'a cause' is not driven purely by policy but also by principle. However although the identification of causal connection is based on principle it is still an evaluative process and as such it ought not be considered part of factual causation.

As Stapleton has highlighted, when Hart and Honoré address 'causally relevant condition' they say that 'a cause is basically like an element in a recipe' but when they discuss 'causal connection' they say that 'a cause is...an intervention in the existing or expected course of events'.¹⁶ To say that both of these inquiries are causal would create confusion whenever the word 'cause' is used as to which of these senses is intended. As Stapleton argues, '[n]ot only are there two distinct underlying enquiries, one historical and one purposive, but the disputes to which they give rise are of a fundamentally different nature'.¹⁷ Glanville Williams explained that Hart and Honoré's contention that ' "it is the plain man's notions of causation (and not the philosopher's or the scientist's) with which the law is concerned"...may be partly true of the legal notion of proximity, but it is not true of the lawyer's use of the notion of but-for causation'.¹⁸ The first stage of identifying 'causally relevant conditions' is the factual aspect of causation which identifies the defendant's negligence as *a cause* and it is from this that Wright later developed the

¹⁴ ibid 305-307.

¹⁵ Wright, 'Causation' (n2) 1739.

¹⁶ Hart and Honoré (n7) 31 and 29. See Jane Stapleton, 'Unpacking Causation' in Peter Cane and John Gardner (eds) *Relating to Responsibility: Essays in Honour of Tony Honoré on his 80th birthday* (Hart Publishing 2001) 159.

¹⁷ Stapleton (n16) 159.

¹⁸ Glanville Williams, 'Causation in the Law' [1961] CLJ 62, 65.

NESS account of causation. It is for this stage that the label of ‘factual causation’ ought to be reserved.¹⁹ The later stages of causal connection and remoteness are both evaluative and address the question of whether the defendant can be considered *the cause* and should be grouped together under the rubric of ‘legal causation’.

Stapleton has highlighted an advantage of limiting the doctrine of causation to the purely factual question of involvement:

This would then require us to locate the normative controversies about their different degrees of *responsibility* for the death under analytical labels such as “duty,” “breach,” ...and so on. The great attraction of this approach is that under these analytical labels, unlike the label of “causation,” it has traditionally been unacceptable merely to assert a conclusion on the basis of “intuition” or “common sense.”²⁰

So although the notion of a causally relevant condition is very inclusive when viewed in isolation, causation alone is not determinative of liability and the other doctrines enable the law to narrow liability from causal responsibility to interpersonal responsibility.

1.2 The robustness of the philosophical account

The NESS test for factual causation that was developed by Wright, building on the work of Hart and Honoré on causally relevant conditions, is based on the dominant philosophical account of causation. This is David Hume’s regularity account of causation and counterfactual analysis, later modified by John Stuart Mill.

Lord Hoffmann has questioned why academics insist on strict factual causation and say that if a court holds something that ‘did not qualify as ‘cause in fact’ as nevertheless satisfying its causal

¹⁹ Stapleton has argued that since ‘legal causation’ is evaluative and not truly a question of causation it should be relabelled ‘scope of liability for consequences’ to avoid confusion (‘Cause-in-Fact and the Scope of Liability for Consequences’ (2003) 119 LQR 388). Analysis of this proposal is outside the scope of this thesis, what matters is that the doctrine of ‘factual causation’ has been defined.

²⁰ Stapleton, ‘Choosing’ (n1) 446.

requirements, then it should be regarded as deeming something to be a cause when it was not *really* a cause'.²¹ He argues:

It is this concept of something having to be *really* a cause according to criteria lying outside the law which puzzles lawyers. On what basis are academic writers entitled to say that judges should take into account a philosophically privileged form of causation which satisfies criteria not required by the law?²²

However, given that causation is essential to corrective justice it is important that the concept of causation adopted in law reflects the natural relation. It is the task of philosophers of causation to understand and explain this, rather than it being a legal creation or fiction. As Moore explains:

[C]orrective-justice...demands a robustly metaphysical interpretation of cause. For legal liability tracks moral responsibility on this view, and moral responsibility is for those harms we *cause*. 'Cause' has to mean what we mean when we assign moral responsibility *for* some harm, and what we mean in morality is to name a causal relation that is natural and *not* of the law's creation.²³

Seeking an explanation from philosophy has been called 'philosophically naïve' by Broadbent who accuses Richard Wright of 'underestimat[ing] how difficult it is to provide a satisfactory account of causation itself'.²⁴ As Fumerton and Kress have also observed, '[p]hilosophers have labored long and hard on the question of how to analyze causation, with a striking lack of success'.²⁵ Furthermore, in the absence of a 'fully adequate universal test for factual causation', Broadbent asserts that there can be no such thing as 'factual causation' and that it will always be determined by the law.²⁶ However while the philosophical account may remain incomplete in places, it is the most complete account of causation available so it provides the law with the most

²¹ Lord Hoffmann, 'Causation' in Richard Goldberg (ed.) *Perspectives on Causation* (Hart Publishing 2011) 5.

²² ibid.

²³ Michael S. Moore, *Causation and Responsibility: An Essay in Law, Morals, and Metaphysics* (Oxford University Press 2009) 95.

²⁴ Alex Broadbent, 'Fact and Law in the Causal Inquiry' (2009) 15 Legal Theory 173, 176.

²⁵ Richard Fumerton and Ken Kress, 'Causation and the Law: Preemption, Lawful Sufficiency, and Causal Sufficiency' (2001) 64 Law and Contemp Prob 83, 102.

²⁶ Broadbent (n24) 178.

robust account of causation that is currently possible. Pragmatism is needed in law because the need to reach concrete decisions precludes waiting for philosophical certainty. Miller explains, '[p]oliticians are not philosophers; they cannot wait for certainty before, for example, banning a drug which appears to have a harmful side-effect. Certainty will be similarly absent in evidence presented in any civil action taken by those so harmed'.²⁷ While there do remain some points of philosophical debate about the meaning of causation, there is a dominant account and it is preferable for the law to adopt this account rather than seeking a distinct definition.

Having separated issues of responsibility, principled and policy-based, from issues of causation, it is now necessary to consider exactly what causation is. What does it mean to be a factual cause, or in Hart and Honoré's classification a 'causally relevant condition'?

2. The philosophical account of causation

To understand what it means to be a cause Hart and Honoré, and subsequently Wright, turned to the dominant philosophical account of causation provided by Hume and Mill.

2.1 Hume: contiguity, priority, constant conjunction and necessity

According to Hume's account of causation the qualities that can be observed in the relationship between causes and effects are contiguity, succession and constant conjunction. These observations form the regularity account of causation. They can be illustrated with an example provided by Hume:

Here is a billiard-ball lying on the table, and another ball moving towards it with rapidity. They strike; and the ball, which was formerly at rest, now acquires a motion...There was no interval betwixt the shock and the motion. *Contiguity* in time and place is therefore a requisite circumstance to the operation of all causes. 'Tis evident likewise, that the motion, which was the cause, is prior to the motion, which was the effect. *Priority* in time,

²⁷ Chris Miller, 'Causation in personal injury: legal or epidemiological common sense?' (2006) 26 LS 544, 547.

is therefore another requisite circumstance in every cause. But this is not all. Let us try any other balls of the same kind in a like situation, and we shall always find, that the impulse of the one produces motion in the other. Here, therefore, is a *third* circumstance, *viz.* that of a *constant conjunction* betwixt the cause and effect. Every object like the cause, produces always some object like the effect.²⁸

Hume then went on to consider the question of whether there is a relationship of necessity between cause and effect, and this is where the novelty of his account lay as Wright explains:

Hume revolutionized philosophic thinking on causation when he insisted that, contrary to the then-popular belief, singular causal judgments are not based on direct perception of causal qualities or forces inherent in objects or events: no such quality or force has ever been identified. Instead, causal judgments are based on the belief that a certain succession of events instantiates one or more causal laws.²⁹

In other words, we observe that necessity is what distinguishes a causal relationship but it does not exist as a quality internal to objects. So for any relationship of objects displaying contiguity, succession and constant conjunction, the element that distinguishes the causal from the non-causal is necessity. This introduces the element of counterfactual analysis in causation that allows us to test whether a factor that satisfies the regularity account is actually a cause. For example, an ‘epiphenomenon’ displays contiguity, succession and constant conjunction but does not cause the outcome: if *c* is an epiphenomenon of the causal history of *e* it is ‘a more or less inefficacious effect of some genuine cause of *e*'.³⁰ An illustration of this is the needle on a barometer dropping before a storm. The drop in pressure *c*, first causes the barometer needle to fall *effect e*, and then causes a storm *effect f*: ‘*c* causes first *e* and then *f*, but *e* does not cause *f*.'³¹ If the barometer is

²⁸ David Hume, *An Abstract of a Treatise of Human Nature*, ed. JM Keynes and P Sraffa (Cambridge University Press, Cambridge 1938) 11f.

²⁹ Wright, ‘Pruning’ (n4) 1019.

³⁰ David Lewis, ‘Causation’ (1973) 70 Journal of Philosophy 556, 557.

³¹ ibid 566.

broken and the needle does not fall this does not affect the fact that the drop in the pressure has occurred so effect, the storm, will still occur.

The necessity requirement explains an aspect of the House of Lords' decision in *Rothwell v Chemical and Insulating Co Ltd*³² concerning claims for pleural plaques. Among other heads of damages, the claimants were unable to recover for the increased risk of contracting asbestos-related illness because pleural plaques do not increase this risk. Pleural plaques will often be observed prior to the development of other asbestos-related illness, however they do not increase the risk of such illness but merely signal that the patient is at an increased risk of contracting an illness because he has inhaled asbestos dust. It is the quality of necessity that differentiates asbestos dust as causative of certain cancers and pleural plaques as non-causative but merely indicative of the risk. Pleural plaques are, in effect, a human equivalent of the needle on a barometer dropping before a storm.

In summary, causation thus consists of three empirical elements: contiguity, succession and constant conjunction; and a non-empirical element: belief in necessary connection.³³ Despite its non-empirical character, Hart and Honoré have observed that necessity has entered modern thinking in an altered form having 'changed its psychological form for a logical one'.³⁴

2.2 Mill: the addition of 'plurality' and 'complexity'

Hume, however, still largely wrote about 'an object' or 'an event' as cause and effect, so of a cause having a single effect, and an effect having a single cause. Mill subsequently made two crucial modifications to Hume's account of causation, recognising the 'plurality of causes' and the 'complexity of causes'.³⁵

³² *Rothwell v Chemical and Insulating Co Ltd and another, Topping v Benchtown Ltd (formerly Jones Bros Preston Ltd), Johnston v NEI International Combustion Ltd, Grieves v F T Everard & sons Ltd and another* [2007] UKHL 39, [2008] 1 AC 281.

³³ Tom L Beauchamp & Alexander Rosenberg, *Hume and the Problem of Causation* (Oxford University Press 1981) 4.

³⁴ Hart and Honoré (n7) 14.

³⁵ Hart and Honoré (n7) 19. Writing about Mill, *System of Logic*.

2.2.1 Complexity of causes

The complexity of causes means that an occurrence does not have a single object or event as its cause, but the cause is the sum total of a set of positive and negative conditions. Drawing on the scenario used by Hume above, the causal law concerning the movement of the billiard ball did not just contain the event of being struck by the other ball, but of being struck by a ball ‘of the same kind in a like situation’. Being struck by a smaller ball, at a different speed, or on a different surface, may not lead to the same result. So an event is caused by the coming together of a set of conditions. Wright explains:

A fully specified causal law or generalization would state an invariable connection between the cause and the consequence: given the actual existence of the fully specified set of antecedent conditions, the consequence must follow.³⁶

Each of the conditions that makes up the complete set has an equal status. Each is equally deserving of being called a cause, because in the absence of any of them the outcome would not have occurred: ‘it is in fact related to the effect in precisely the same way as the other constituents of the set’.³⁷ The idea that every factor that makes up the set is equally deserving of the label ‘cause’ is a situation Mackie describes as being ‘symmetrical with respect to all the factors’.³⁸ Causation, in Honoré’s words, provides a ‘recipe’ for the set of conditions that are together sufficient to produce a given result.³⁹

As noted in section 1.1, the causal inquiry in negligence is focused on the question of whether the negligence was a cause of the damage, so it is unnecessary to provide the whole ‘recipe’. This does not prevent the legal concept of causation from being based on the philosophical account since it is possible to remain faithful to Mill’s explanation of causation whilst on a practical level

³⁶ Wright, ‘Causation’ (n2) 1789.

³⁷ Hart and Honoré (n7) 21.

³⁸ Mackie (n10) 34.

³⁹ Tony Honoré, ‘Necessary and Sufficient Conditions in Tort Law’ in David G Owen (ed.) *Philosophical Foundations of Tort Law* (Oxford University Press 1995) 375.

being content to only partially populate the sufficient set that instantiates the causal law in question as Wright explains:

In the typical singular causal statement, the causal assertion includes, explicitly or implicitly, only a few of the antecedent conditions but nevertheless asserts that they were only part of an incompletely specified (and incompletely understood or known) set of actual conditions that was sufficient for the occurrence of the consequence.⁴⁰

Beever criticises this position, arguing that it means that ‘the law’s understanding of causation is looser than science’s’ and that this is inconsistent with the requirement that ‘the lawyer’s conception of causation must be philosophical and scientific for it to be a conception of causation’.⁴¹ Beever provides an example of a barn that is burnt down because the defendant placed his rick of hay against it. He says that science would require not just the presence of the rick of hay, but all the other conditions making up the sufficiency of the set for the causal law, such as the dryness of the hay, sufficient heat, the barn being made of flammable material, the presence of oxygen etc. In contrast, Wright would say that these conditions are not relevant in law, because the law is satisfied to say that the rick of hay was one antecedent condition which was part of an incompletely specified set. This leads Beever to believe that Wright is saying that the legal concept of causation is different to the scientific concept of causation. To support this argument he says that the other constituents of the set are not irrelevant because the court would not be likely to conclude that the rick of hay caused the destruction of the barn if the rick was soaking wet, or the hay did not ignite, or the barn was made of concrete. Whilst Beever is right to say that the court would reach a different conclusion about causation if it knew that any of these other conditions were present, this does not mean that it adopts a different, unscientific and unphilosophical concept of causation if it does not list all of the conditions. The underlying concept of causation is the same as the philosophical or scientific one, but the court’s purpose is not to discover every element of the causal set. Instead the court’s purpose is to determine

⁴⁰ Wright ‘Causation’ (n2) 1789.

⁴¹ Allan Beever ‘Cause-in-fact: Two Steps Out of the Mire’ (2001) 51 UTLJ 327, 344.

simply whether one specific candidate condition, the defendant's negligence, in fact belonged to the set. The court therefore relies on the information presented to it, so if the defendant established that counter-conditions, such as the rick being soaking wet, had been present, then the court will add these into the causal set. But if no such conditions are presented, then as a practical measure the court is justified in being satisfied that the rick was a cause of the destruction of the claimant's barn. The concept of causation does not change, but pragmatism leads the law to be satisfied to address only those candidate conditions that are presented to it.

2.2.2 Plurality of causes

In addition to causes being 'complex', they are also 'plural': there is not just one set of conditions that is capable of causing the outcome, instead there are a number of different possible sets, each of which is capable of causing the same outcome. So, for example, a billiard ball may move because it is struck by another ball, or because the table is tilted or because air is blown at it etc. In summary, a single outcome is caused by a set of conditions rather than having a single cause, and various different sets of conditions can exist that are all capable of causing the same single outcome.

In this account necessity is still essential as the quality that distinguishes causally relevant antecedents:

[O]ne must be careful when constructing a causal law or generalization to distinguish the causally relevant antecedent conditions from the causally irrelevant antecedent conditions. This differentiation is necessary to insure that the set of jointly sufficient antecedent conditions includes only those that are indeed invariably connected with the consequence. Thus, the antecedent conditions must be restricted to those that are *necessary* for the sufficiency of the set.⁴²

⁴² Wright, 'Causation' (n2) 1790.

Necessity has been confined to a smaller role because for a given effect it is not *necessary* for it to be preceded by a particular cause since there exist a number of possible sets capable of causing the effect. So although each condition is necessary for the sufficiency of a particular set, each particular set can only be described as ‘sufficient’ to bring about the effect.

Hart and Honoré noted a philosophical objection to the idea of plurality of causes in the argument that an equal plurality of effects must exist but we fail to differentiate effects with the same precision that we apply to causes: the ‘apparent plurality of causes is due to our failure or inability to analyse the effects with the same particularity as the causes’.⁴³ If we were able to particularise any single effect sufficiently we would see that it only has one possible cause. This is similar to saying that we must be more precise than simply saying a person died and, for example, specify that he died from a heart attack, or from poisoning. It is simply saying that we need to continue this process of precision in our description of the event until we have as many effects as we have causes.

They noted that this philosophical objection is not universally accepted, but argued that whether or not it is well-founded, ‘the lawyer, the historian, and the plain man accept the doctrine that an event of a given kind may be produced by different causes’.⁴⁴ That the doctrine of plurality of causes is accepted can be illustrated by close consideration, for example, of the cause of a heart attack. A heart attack occurs because part of the heart muscle dies when it is starved of oxygen, usually because of a blockage in a coronary artery. Such blockages can occur for a multitude of reasons with risk factors including smoking, high blood pressure, lack of exercise, and poor diet.⁴⁵ Rather than attempting to analyse the effect, the heart attack, in ever more detail, the ordinary man accepts that the effect has a plurality of possible causal sets, each containing one of the many possible combinations of these, and other, risk factors. Hart and Honoré’s pragmatism is justifiable on the basis that if we did not accept plurality of causes then we would never accept

⁴³ Hart and Honoré (n7) 20.

⁴⁴ Hart and Honoré (n7) 20.

⁴⁵ <http://www.nhs.uk/Conditions/Heart-attack/Pages/Causes.aspx> Accessed 14th January 2010.

any causal law as being true, and never be able to make causal judgments. Causal conclusions cannot wait until the moment when, indeed if, science advances to a point where effects can be adequately particularised. In other words, this is one point where pragmatism is necessary since the law cannot wait for philosophy and science to become settled.⁴⁶

3. Tests for causation

So far this chapter has identified the demands that negligence places on the doctrine of causation, and the philosophical account of the concept of causation. It is now possible to build on these foundations to analyse the possible tests for causation. It will be argued that the NESS test should be adopted instead of the but-for test because it reflects the concept of causation more accurately. While the but-for test has the attraction of apparent simplicity compared to the NESS test, it only functions correctly in straightforward cases, and in those cases the NESS test operates just as simply. In the more complex cases of over-determined causation, the but-for test fails because it is not an accurate test of causation, so courts have developed exceptions whose scope and meaning are unclear. In contrast, the NESS test is able to cope with these more complex cases, and because it accurately reflects the concept of causation it enables us to visualise the causal problems and solutions without resorting to exceptional tests.

3.1 Complexity and plurality of causes as necessary and sufficient conditions

The philosophical account of causation described above established the following qualities as characteristics of causes: contiguity of cause and effect, priority in time and constant conjunction. In addition, necessity is the quality that distinguishes causal factors from non-causal factors. Moreover, causes are complex: an outcome is the result of a set of conditions coming together; and plural: multiple possible sets can each bring about the same outcome. The NESS

⁴⁶ See text to (n27).

test will be shown to be preferable to the but-for test on a theoretical level because it more accurately reflects these qualities.

There are different degrees to which a single factor, such as a defendant's negligence, might be thought sufficient or necessary for the occurrence of the harm. These are explained by Wright:

In descending order of stringency, a strict-necessity test requires that Q be necessary for the occurrence of R whenever R occurs; a less stringent, strong-necessity test requires only that Q have been necessary for the occurrence of R on the particular occasion, considering the circumstances that existed on the particular occasion; and the least stringent, weak-necessity test requires only that Q have been a necessary element of some set of actual conditions that was sufficient for the occurrence of R ... A strict sufficiency test requires that Q be sufficient by itself for the occurrence of R ; a less stringent, strong-sufficiency test requires only that Q be a necessary element of some set of existing conditions that was sufficient for the occurrence of R ...; and the least stringent, weak-sufficiency test "requires" only that Q be a part of some set of existing conditions that was sufficient for the occurrence of R .⁴⁷

Some of these types of test can easily be discarded because they clearly do not reflect the characteristics of causes. Strict necessity would mean that a particular effect only ever has one possible cause, but this is precluded by the notion of plurality of causes. Similarly, strict sufficiency, the idea that one single factor can be sufficient to bring about an effect, is precluded since causes are 'complex'. Weak sufficiency does not incorporate the vital characteristic of necessity to any degree so it is unable to determine whether a factor was a cause. The remaining possibilities are strong necessity and weak necessity (or its equivalent, strong sufficiency). The but-for test asks whether a factor was necessary for the occurrence of harm on the particular occasion so it is a test of strong necessity. The NESS test does not require the factor to have been necessary for the harm on the particular occasion, but allows for the possibility that more

⁴⁷ Wright 'Pruning' (n4) 1020.

than one sufficient set of factors may have existed on the particular occasion. As long as the factor was necessary for the sufficiency of one of the sets that actually occurred on the particular occasion then it was a cause. The NESS test therefore corresponds to weak necessity.⁴⁸ Although the but-for test acknowledges the plurality of causes insofar as it recognises that a variety of different sets *could* potentially cause a particular outcome, it does not take the notion of plurality to its logical conclusion because it assumes that each potential set only ever occurs singularly.⁴⁹ The NESS test recognises that more than one set may *actually* occur simultaneously. Practical examples will show that this complete incorporation of plurality is correct.

3.2 But-for and NESS in simple cases

The but-for test is illustrated clearly by the case of *Barnett v Chelsea and Kensington Hospital Management Committee*⁵⁰ where the claimant's husband became ill after drinking a cup of tea. He went to hospital where the staff were negligent in their failure to examine and treat him, instead sending him home with advice to see his GP the next day. The victim died from arsenic poisoning in his tea but despite the negligence of the hospital staff his claim failed at the causation stage because but-for their negligence he would still have died. Given the time that had elapsed between drinking the tea and arriving at hospital, the time it would have taken to examine and diagnose the illness and begin treatment, the treatment would have failed to save him. Given the circumstances, including the stage of advancement of the poisoning, the doctor's negligence was not necessary for the outcome, it made no difference. Herein lies the appeal of the but-for test, it answers the question 'did the defendant's act make a difference?'.⁵¹ Yet it is widely recognised that the but-for test is of limited use because as well as excluding irrelevant

⁴⁸ Wright, 'Pruning' (n4) 1021

⁴⁹ Mackie's INUS test ('insufficient but necessary part of an unnecessary but sufficient set' test) incorporates the plurality of causes since it recognises that a particular sufficient set of conditions need not be necessary for the outcome. See JL Mackie, *The Cement of the Universe: A Study of Causation* (Clarendon Press 1974). For any particular instance of causation, however, it only recognises the occurrence of one sufficient set which, as Wright explains, converts the INUS account into the *sine qua non* ('but-for') account. See Wright, 'The NESS Account' (n3) 288. Since it is applied in the same way as the but-for test it will not be analysed separately.

⁵⁰ [1968] 2 WLR 422 (QBD).

⁵¹ Moore (n23) 84.

conditions ‘it can also exclude others that are relevant’⁵² in cases where an outcome is ‘over-determined’ as in the classic double-hit hunters scenario discussed in the next section.

The NESS test was developed by Richard Wright in response to the weaknesses of the but-for test and built upon ideas initially developed by Hart and Honoré in *Causation in the Law*. In this account of causation something is a cause if it was a Necessary Element of a Sufficient Set: ‘*a particular condition was a cause of (condition contributing to) a specific consequence if and only if it was a necessary element of a set of antecedent actual conditions that was sufficient for the occurrence of the consequence.*’ (Note that the phrase “*a set*” permits a plurality of sufficient sets.)⁵³ The first part of this is uncontroversial, since causes are ‘complex’ any outcome is caused by the coming together of a set of conditions which are together sufficient to cause it. The novelty of the NESS test is that it fully incorporates plurality, so while in most instances one particular outcome will only have been preceded by one of these possible sets, the NESS test recognises that multiple sets may actually occur.

Where only one set actually occurs the NESS test would not complicate the way the law currently works. For example in *Barnett* there are many ways that the claimant’s husband’s death could have been caused but only one, poisoning with arsenic, actually occurred. The doctor’s negligence was not necessary for the sufficiency of the set of conditions that actually occurred because the treatment would still have been too late even if the doctor had examined the patient. In these straightforward cases the NESS test would effectively collapse into the but-for test; the but-for test is shorthand for the NESS test where there is only one set of conditions. This makes it clear that widespread adoption of the NESS test would not have the effect of complicating cases generally which should allay possible fears that adopting the NESS test would lead to increased litigation.

⁵² Honoré (n39) 363-4.

⁵³ Wright ‘Causation’ (n2) 1790.

Since strong necessity (but-for) is more stringent, a factor that is strongly necessary will also satisfy a test of weak-necessity. But the reverse is not true – a factor that fails the test of strong-necessity may still be a cause because causation only requires weak-necessity. The but-for test can tell us that something *was* a cause, it is ‘a test of inclusion’,⁵⁴ but it cannot accurately tell us whether something *was not* a cause. As Honoré has said:

No one will deny that the but-for test has in many instances a heuristic value: it often provides a quick way of testing the existence of causal connection. However, it is another matter whether it is part of the *meaning* of ‘causally relevant condition’ or ‘cause’.⁵⁵

The but-for test fails to deal adequately with more complex causal problems because it does not accurately correspond to the meaning of causation. So the fact that we are often able to rely on the but-for test should not be allowed to mislead into the belief that it is the best test to use. As Wright argues, ‘the substitute must give way to the more accurate and comprehensive concept when the situation is more subtle and complex’.⁵⁶

3.3 Over-determined causation

‘Over-determination’ means that there seem to be too many causes. Where this takes the form of ‘duplication’ two or more possible causes occur at the same time, and in ‘pre-emption’ the possible causes occur one after the other.⁵⁷

⁵⁴ Wright, ‘Pruning’ (n4) 1022.

⁵⁵ Honoré (n39) 367.

⁵⁶ Wright, ‘Causation’ (n2) 1792.

⁵⁷ Hart and Honoré used different labels to describe the forms of over-determined causation, stating that ‘an event may be causally overdetermined either because two conditions each sufficient though not necessary in the circumstances for its occurrence are present together (additional causes) or because they are so related that if one had not been present the other would have been (alternative causes)’ (Hart and Honoré (n7) xl). Wright’s terminology of ‘duplication’ and ‘pre-emption’ is adopted here for two reasons. First, given the emphasis on the importance of clear use of technical terminology, it is preferable to avoid the term ‘alternative causes’ which is sometimes used with a different meaning in the caselaw as a contrast with ‘cumulative causes’. Secondly, Wright’s account benefits from greater analytical clarity on this point because, as he observed, ‘Hart and Honoré also submerged and sometimes confused the critical distinction between duplicative and pre-emptive causation by constructing an overlapping typology of overdetermined cases’, so they did not always maintain a clear distinction between the two forms of over-determination (Wright, ‘The NESS Account’ (n3) 287).

3.3.1 But-for, NESS and pre-emption

Pre-emption describes a scenario where two acts could each have been sufficient to cause the harm, but one occurs before the other so the second one never gets to run its course. This is illustrated by the classic desert traveller scenario: *A* puts poison in the water keg of a desert traveller which is sufficient to kill him if he drinks it but before he drinks the water, *B* empties the keg, and the traveller dies of thirst. The but-for test leads to the conclusion that neither act was a cause of the traveller's death: if he had not died of thirst he would have died from poisoning and *vice versa*. Effectively, the but-for test 'produces the absurd result that the plaintiff's injury was uncaused'.⁵⁸

It has been suggested that the but-for test would identify *B* as the cause of the desert traveller's death if the outcome was described more precisely as 'death when and as it happened' i.e. death through dehydration at that particular moment. This tactic of 'detailing the injury', Wright argues, is 'nothing more than proof by tautology' because '[t]he factors believed to be causally relevant...are incorporated into the description of the manner of occurrence of the injury..., and they are then demonstrated to be causally relevant because we cannot construct that precise description without them'.⁵⁹ It is therefore clear that the but-for test cannot cope adequately with pre-emptive causation.

The legal solution to this absurd outcome would be to impose liability on one or both parties despite the failure to satisfy the legal test for causation. Such a solution is driven by the instinctive knowledge that the death cannot be 'uncaused', one or both parties must be responsible for the death despite the difficulty of establishing a causal link. Thus reliance on the but-for test effectively turns the legal solution to pre-emption into a value judgment as to responsibility. But, as Hamer has argued, 'the over-determination problem is one of fact and

⁵⁸ Beever (n41) 331.

⁵⁹ Wright, 'Causation' (n2) 1778.

philosophy rather than value and policy'.⁶⁰ This is also an unnecessarily complicated ‘solution’ because there is not really a ‘problem’ regarding the existence of a causal link and this can be seen by explaining and applying the NESS test.

In contrast to the but-for test, the NESS test asks whether the negligence was a ‘necessary element of a set of antecedent *actual* conditions that was sufficient for the occurrence of the consequence’.⁶¹ In the desert traveller scenario, by poisoning the water flask *A* created the *potential* for a set to occur that would be sufficient to bring about the traveller’s death. However, for the set of conditions to be complete, in other words for it to *actually* be sufficient, the traveller must drink the water containing the poison. Until he drinks the poison there is not an *actually occurring* set of conditions sufficient to cause his death that contains *A*’s act of adding poison to the water. Since *B* then emptied the water (and therefore also the poison) the traveller never actually drank the poison. This means that *A*’s act was prevented from becoming part of an *actually occurring* set of conditions that were together sufficient to cause the death, so *A* is not a cause of the harm.⁶² A set of conditions jointly sufficient to cause the death now exists containing just *B*’s act of emptying the flask. This set of conditions runs its course so it is an *actually occurring* set, so *B* is a cause.

3.3.2 But-for, NESS and duplication

The plurality of causes means that one outcome could be caused by a variety of sufficient sets and usually only one of these sets will actually have occurred, but over-determination in the form of duplication arises when more than one sufficient set of conditions actually occurs. Duplication, in its simplest form, is illustrated by the classic ‘double-hit hunters’ example: Hunter *A* fires his gun and the bullet hits a walker. The walker dies and the bullet was sufficient to cause

⁶⁰ David Hamer, ‘Chance would be a fine thing? Proof of Causation and Quantum in an Unpredictable World’ (1999) 23 Melb U L Rev 557, 571.

⁶¹ Wright, ‘Causation’ (n2) 1790 (emphasis added).

⁶² See Wright, ‘Causation’ (n2) 1795.

his death. In the ordinary course of events, Hunter *A* would be considered to have caused the death because his bullet was sufficient to result in the death. However, Hunter *A*'s bullet was 'duplicated' by Hunter *B* simultaneously firing his gun and also hitting the walker with a bullet that was sufficient to cause the death.

If we apply the but-for test to determine whether Hunter *A* was a cause of the walker's death then the answer is 'no'. But-for Hunter *A*'s shot, the walker would still have died because Hunter *B*'s bullet was sufficient to kill him, so Hunter *A* made no difference to the outcome and was not a cause of the death. Likewise, but-for Hunter *B*'s shot, the walker would still have been killed by Hunter *A* so Hunter *B* made no difference to the outcome and is not a cause. Once again, the but-for test 'produces the absurd result that the plaintiff's injury was uncaused'.⁶³

In contrast, the NESS test fully incorporates the plurality of causes and asks whether the negligence was necessary for the sufficiency of *a* sufficient set that actually occurred.⁶⁴ It therefore allows for the existence of more than one actually occurring set. If we subtract Hunter *B*'s shot we are left with a set of conditions containing Hunter *A*'s shot, along with wind speed and direction, the walker's presence, the walker still being alive etc., that are sufficient to cause the walker's death, and Hunter *A*'s shot is necessary for this set to be sufficient. Likewise if we subtract Hunter *A*'s shot we are left with a set of conditions containing Hunter *B*'s shot that is sufficient to cause the walker's death and Hunter *B*'s shot is necessary for the sufficiency of this set. The plurality of causes, as identified by Mill, means that there are also many other possible sets that would hypothetically have been sufficient to cause the walker's death had they occurred at that moment, for example a tree toppling onto him, but none of these other sets *actually* occurred. However, the fact that two sufficient sets independently and simultaneously have *actually* occurred does not prevent either set from being a cause. Usually only one set will actually occur, but on the rare occasion that two sets *actually* occur then both are causes.

⁶³ Beever (n41) 331.

⁶⁴ See text to (n61).

Whereas the but-for test told us that Hunter A *made no difference to the outcome* because the walker would have died anyway so Hunter A is not a cause, the NESS test tells us that in order for the walker's death to have been prevented Hunter A's shot *would have to be absent* so Hunter A is a cause, and that Hunter B's shot would also have to be absent so Hunter B is also a cause.⁶⁵ This is a subtle yet significant shift in emphasis and one which better reflects the reality that given the plurality of causes it is possible, albeit rare, for two or more causes independently to occur at the same time.

It is clear that the NESS test, rather than the but-for test, accurately reflects and tests the qualities that were shown by Hume and Mill to characterise causes. The NESS test has therefore been described as 'a real contribution to legal analysis. It provides an extremely helpful way of conceptualizing the nature of causal problems, and it offers a rational process for identifying causes in over-determined cause cases'.⁶⁶ This means that on a theoretical level it would be preferable to adopt the NESS test. It represents a subtle shift in emphasis from the but-for test which treats as a cause 'something in the absence of which harm would not have come about', to calling a cause 'something such that in the prevailing conditions harm comes about' under the NESS test.⁶⁷ This may seem subtle but it is very significant because it accurately translates the meaning of causation and the characteristics of causes into a workable test. In practice the NESS test will frequently collapse into the but-for test because most cases involve straightforward causal models, but NESS is to be preferred because it is able to tackle complex issues of causation in a way that the but-for test cannot. This choice is 'not governed by policy considerations, but rather by how well each test corresponds with our intuitive concept of causation' based on the philosophical explanation of causation.⁶⁸

⁶⁵ Stapleton 'Choosing' (n1) 438.

⁶⁶ David A Fischer, 'Insufficient Causes' (2005) 94 Ky LJ 277, 281.

⁶⁷ Honoré (n39) 385.

⁶⁸ Wright, 'Pruning' (n8) 1020.

4. Overcoming potential problems with NESS

The final section of this chapter will use the NESS test to overcome some common legal problems that are currently addressed using the *Wardlaw* test of ‘material contribution to harm’,⁶⁹ but first the penultimate section addresses some potential problems with the NESS test. The problem with the *Wardlaw* test is that it is not clear how exceptional it is; it may be an exception to the but-for test that compensates for the conceptual inadequacies of that test, or it may involve a relaxation of the causation requirement itself. This problem arises because it is unclear precisely what causal problem is addressed by the test. The NESS test would similarly face potential problems in more complex cases if causation were analysed in isolation from the negligence inquiry. This section therefore uses the damage doctrine to incorporate a more nuanced idea of ‘outcome’ into Wright’s NESS test. In the final section it will then be seen that the NESS test allows us a level of conceptual clarity that is absent in the *Wardlaw* test.

Wright has explained that duplicated causation can take a more complex form than that encapsulated in the double-hit hunters scenario. The double-hit hunters scenario involves ‘symmetrical duplication’ since each shot is part of a set of conditions that is sufficient for the walker’s death independently of the other shot. Duplication is ‘asymmetrical’ where there are various sources of the harmful agent and the total dose exceeds a sufficient amount (hence causation is over-determined) but each individual source is dependent on other sources to make a sufficient set. Since each source is not independently sufficient but is dependent on other sources, the duplication is said to be ‘asymmetrical’. Wright explains this problem more fully using the example of river pollution, which will be explained below. He explains that the NESS test can be used to address asymmetrical duplication, but his application of the NESS test seems to be so inclusive that it risks becoming useless as a functioning test for causation in negligence. It is important, however, to apply the NESS test within the context of the negligence inquiry, and defining the damage precisely at the outset shows that asymmetrical duplication is actually a

⁶⁹ *Bonnington Castings v Wardlaw* [1956] AC 613 (HL).

very limited causal problem. In this limited problem, the NESS test is very inclusive but is also accurate. Again it is important to remember that causation is not determinative of liability; causation is a factual relation so the contours of the test for it cannot be driven by the perceived fairness of the outcome. The limits of liability must be addressed by other doctrines, and here there are issues relating to quantification and to the effects of joint and several liability.

4.1 Defining key causal terms

A distinction is drawn in negligence cases between so-called ‘divisible’ and ‘indivisible’ damage. This terminology broadly corresponds to the scientific terminology of ‘stochastic’ or ‘non-stochastic’ disease. An indivisible, or stochastic, disease is an ‘all-or-nothing’ phenomenon:

There is almost always a clear distinction between people who have the disorder and those who do not. The disease can differ in its severity from one case to another, but the factors that determine its severity are in general quite different from those that determine whether or not it occurs at all.⁷⁰

In contrast, non-stochastic diseases ‘occur in a continuum of severity with no clear distinction between cases and non-cases’ although ‘[d]octors may adopt a dichotomous case definition to facilitate the practical management of patients’.⁷¹ In other words, the disease gradually worsens and at a particular level of severity it is deemed to exist although ‘the chosen threshold is to some extent arbitrary’.⁷² Divisible disease is therefore dose-related and each exposure to the causal agent makes the illness more severe. For example in *Thompson v Smiths Shiprepairs*,⁷³ the claimant’s loss of hearing was divisible because it gradually worsened with exposure to noise. Similarly the claimant’s pneumoconiosis in *Cartledge v E Jopling*,⁷⁴ and the claimant’s asbestosis in

⁷⁰ David Coggon and Anthony Newman Taylor, ‘Causation and attribution of disease in personal injury cases: a scientific perspective’ [2009] JPI Law 12, 13.

⁷¹ *ibid.*

⁷² *ibid.*

⁷³ *Thompson v Smiths Shiprepairs (North Shields) Ltd* [1984] QB 405 (QBD).

⁷⁴ *Cartledge v E Jopling & Sons Ltd* [1963] AC 758 (HL).

*Holtby v Brigham & Cowan*⁷⁵ were divisible diseases because they became more severe the more the claimant inhaled silica dust or asbestos respectively.

As Coggan and Taylor explain, the term ‘stochastic’ means relating to chance, so the probability of suffering a stochastic disease depends on exposure to risk factors. This means that although it is not possible to state whether an individual will develop a particular disease, it is possible to say how their probability of developing the disease is affected by particular factors. So ‘the potency of a cause can be quantified in terms of the “relative risk” that it carries for the disease...For example, smoking 20 cigarettes per day for 40 years might increase the risk of lung cancer 15-fold in comparison with a non-smoker – a relative risk of 15’.⁷⁶ In contrast, because non-stochastic diseases gradually develop in severity, we cannot sensibly talk about the probability of developing the disease and instead ‘the potency of a cause is most meaningfully quantified by its average impact on the severity of a disease’.⁷⁷ The term ‘divisible’ is effectively a lay-term to describe the relationship between the disease and its causes in non-stochastic diseases where each exposure to the harmful agent makes the disease worse.

It is important to note that describing the outcome as divisible is not the same as describing the exposure to sources or triggers of the harmful outcome as ‘divisible’ or ‘cumulative’.⁷⁸ If the exposures are ‘divisible’ in the sense that there were a number of exposures, simultaneous or spread over a period of time, the damage may still be indivisible if it occurs once a particular threshold has passed but does not continue to increase in severity with the increase in dose.⁷⁹

⁷⁵ *Holtby v Brigham & Cowan (Hull) Ltd* [2000] 3 All ER 421 (CA).

⁷⁶ Coggan and Taylor (n70) 14.

⁷⁷ *ibid.*

⁷⁸ See also *Sienkiewicz v Greif (UK) Ltd* [2011] UKSC 10, [2011] 2 WLR 523, [12]-[14] (Lord Phillips)

⁷⁹ This will be revisited in relation to terminology used to describe *Wardlaw* since some commentators use the terms alternative and cumulative to signal that the exposures are divisible/indivisible, others use these terms to signal divisibility of damage.

4.2 The problem of asymmetrical duplication

To illustrate asymmetrical duplication, Wright considers a river pollution scenario where five units of pollution were necessary and sufficient for the injury that followed, and each of seven defendants discharged one unit of pollution.⁸⁰ He explains two details in a footnote. First, he ‘assume[s] that the injury was not accelerated or aggravated by the extra units of pollution. If it was, causal contribution would be even clearer’.⁸¹ Second, he assumes ‘that the units of pollution arrived simultaneously at the site of the injury. Obviously, if five units arrived before the other two and produced the injury before the other two arrived, the first five units were causes of the injury and the last two were not. Their potential effects were pre-empted by the effects of the first five’.⁸² The NESS test, as will be explained later, considers each unit to be a cause. These details, however, are immensely significant in understanding the causal problem so they should not be confined to a footnote. The facts must be addressed in more detail before addressing the application of the NESS test to them.

First of all, it is essential to identify whether the injury was aggravated by the extra units of pollution. If each unit aggravated the injury, i.e. made it more severe, then the damage occurs along a continuum of severity and is ‘divisible’.⁸³ As explained in Chapter One, Stapleton insists:

It cannot be over-emphasised that the formulation of the ‘damage’ forming the gist of the action *defines* the causation question. Logically one can only deal with causation after one knows what the damage forming the gist of the action is.⁸⁴

It is essential to identify the damage at the outset, and to determine whether it is divisible or indivisible. If the damage is divisible then each unit of pollution causes a portion of the damage. So where the damage is divisible, the duplication problem simply does not arise because each

⁸⁰ Wright, ‘Causation’ (n2) 1793.

⁸¹ ibid.

⁸² ibid.

⁸³ If it was ‘accelerated’ then the damage is actually still indivisible, this point will be revisited later in relation to indivisible damage.

⁸⁴ Jane Stapleton, ‘The Gist of Negligence: Part 2 The Relationship Between ‘Damage’ and Causation’ (1988) 104 LQR 389, 393.

unit causes the damage to be more severe.⁸⁵ (This does not mean causation in divisible damage cases is entirely unproblematic, indeed it will be analysed in more depth later, it just means that where the damage is divisible we do not face a problem of duplication.)

It should also be noted that the problem of asymmetrical duplication, as the name implies, is unique to duplication rather than pre-emption, as Wright commented in his footnote. If the units of pollution were added consecutively rather than simultaneously then those units that were added after the harm occurred (the death of the fish) would be pre-empted from becoming causes. They could never form part of a causally sufficient set that *actually* occurred.

In practical terms this means that the problem of asymmetrical duplication is confined to very limited circumstances. It only arises where the damage is indivisible and various sources of the harmful agent combine to form an indecipherable mass so that they operate simultaneously.

4.2.1 The solution to asymmetrical duplication

In the river pollution example of asymmetrical duplication, application of the but-for test would lead to the conclusion that none of the individual units was a cause of the injury in the sense of strong necessity because the remaining six units would still have caused it. The but-for test once again suggests that the injury was ‘uncaused’.

Yet the solution under the NESS test is not as simple as it is in the double-hit hunters scenario. In that scenario, the duplication is symmetrical because each shot is independently sufficient to cause the injury. To avoid the injury, the each shot would have to be absent. But where the duplication is asymmetrical as in this river pollution scenario, each unit is not independently sufficient, so we cannot say of any individual unit that it must be absent if the injury is to be avoided. Since no individual unit is either strongly necessary (a but-for cause) or independently sufficient (symmetrical duplication) its presence or absence seems to be irrelevant to the outcome. Yet the same could be said of each of the seven units, returning us to a position where

⁸⁵ Martin Hogg, ‘Developing Causal Doctrine’ in Richard Goldberg (ed) *Perspectives on Causation* (Hart Publishing 2011) 42.

the harm seems to be mysteriously ‘uncaused’, and logically we know this is not true. However, Wright explains that this is not because of a weakness in the NESS test, it is because the NESS test is being misapplied. If the NESS test is applied correctly, each unit is still a cause of the injury:

[E]ach defendant's one unit was necessary for the sufficiency of *a* set of actual antecedent conditions that included only four of the other units, and the sufficiency of this particular set of actual antecedent conditions was not affected by the existence of two additional duplicative units.⁸⁶

This is because of the plurality of causes whereby there can be more than one sufficient set for any outcome, and it is possible for more than one set *actually* to occur. As noted in an earlier section, the philosophical account of causation involves strong sufficiency (or its equivalent, weak necessity) so it does not require *independent* sufficiency. Instead of asking whether a particular unit would have to be absent for the harm to be avoided (strong necessity), the correct question is whether *the set* containing that unit would have to be absent for the harm to be avoided (weak necessity). The situation appears complicated. However, this is only because the two sets are not independently sufficient so there is an overlap between sets.

Wright continues to explain that this conclusion is logically the same whether each of the seven units was contributed by a different individual, or whether one person contributed five units and another contributed two units.⁸⁷ Since the units combine in an indecipherable mass the causal status of a unit of pollution is not dependent on whether the remaining units were contributed by one person or by many people. The two units contributed by the second person are still ‘necessary for the sufficiency of a set of actual antecedent conditions that included only three of the first defendant's five units, a set whose sufficiency was not affected by the existence of two

⁸⁶ Wright, ‘Causation’ (n2) 1793.

⁸⁷ Richard Wright, ‘Once More into the Bramble Bush: Duty, Causal Contribution, and the Extent of Legal Responsibility’ (2001) 54 Vand L Rev 1071, 1106.

additional duplicative units also provided by the first defendant.⁸⁸ The fact that the two sources occur at the same time, or combine to form an indecipherable whole, makes this disaggregation of the five unit contribution into its component units logically possible: if we are able notionally to disaggregate the defendant's contribution from the other contribution then logically we can also notionally disaggregate the other contribution.

Wright thus argues that the same causal situation exists whether the harmful agent is one that would normally be thought of as existing in 'units' or not, so if two or more fires merge to destroy a building, each is a cause of the destruction even if it would have been insufficient on its own, and even if one of the other fires alone would have been sufficient. The second fire 'was necessary for the sufficiency of a set of actual antecedent conditions which included another fire...that was "*at least* large enough to be sufficient for the injury if it merged with a fire the size of the second fire"'.⁸⁹ It will be argued below that in principle Wright is correct to assert that disaggregation is appropriate even when the harmful agent is something such as fire that would not normally be thought of as existing in units. However the practical example that he uses to support this assertion is inappropriate because the harm in cases of fire may be divisible so it should not be used to support an argument concerning an indivisible harm.⁹⁰

Wright has since come to question this 'at least enough' formulation, describing it as 'overly demanding'.⁹¹ This description is surprising given that his 'at least enough' formulation seems to make the NESS test very inclusive because in cases of duplication it allows any contribution, which is neither strongly necessary nor independently sufficient, to be labelled as a cause by constructing a set containing that contribution and just enough of the other contributions. He explains:

⁸⁸ Wright, 'Causation' (n2) 1793.

⁸⁹ *ibid* 1793.

⁹⁰ For example if two fires combine to damage a house it may be the case that one fire would have damaged one room, the second fire another room so the damage would be divisible, but if a person was trapped and killed by the fire their death would be an indivisible damage.

⁹¹ Wright, 'The NESS Account' (n3) 291.

My initial elaborations of the NESS account were overly demanding. I incorporated the weak-necessity requirement in the definition of singular instances of causation. As I have previously stated, this is too restrictive. The weak-necessity requirement is sufficiently incorporated in a properly formulated causal law, which contains in its antecedent only those abstract conditions the instantiation of which is necessary for the sufficiency of the set of conditions that is sufficient for the immediate instantiation of its consequent. When analysing singular instances of causation, an actual condition c was a cause of an actual condition e if and only if c was a part of (rather than being necessary for) the instantiation of one of the abstract conditions in the completely instantiated antecedent of a causal law, the consequent of which was instantiated by e immediately after the complete instantiation of its antecedent, or (as is more often the case) if c is connected to e through a sequence of such instantiations of causal laws. This formulation of the requirement for a NESS condition is more straightforward and simpler to apply than my initial formulations, which requires ‘at least so much’ description of actual conditions in some situations in order to (validly) treat other conditions as NESS conditions.⁹²

This is unconvincing. This seems like a looser approach to causation where the NESS test allows us to understand whether a condition is capable of being a cause but does not allow us to determine whether it actually was a cause on a particular occasion. The NESS ‘test’ thus conceived seems to explain causal generalisations but not to actually test whether something was a cause in a particular instance. This would mean that it has little or no value as a legal test where the law requires a causal link to exist between the particular defendant’s negligence and the particular claimant’s loss.

However, Wright’s previous ‘at least enough’ formulation is appropriate in cases of asymmetrical duplication because asymmetrical duplication only arises where the sources of the harmful agent are factually combined into an indecipherable mass. Since the units are physically combined

⁹² ibid.

together their effects overlap so we are able to construct overlapping sufficient sets of units containing A 's unit and 'just enough' of the other units, or B 's unit and 'just enough' of the other units. It is impossible to say that five particular units had an effect and the remaining units had no effect because they were all present together and all operated at the same time – they formed an indecipherable whole so the effects of each unit are similarly indecipherable.

The 'at least enough' formulation would be unnecessary and inappropriate in other scenarios that do not involve indivisible damage and sources of the harmful agent combining in an indecipherable mass. If only five units of pollution are added to the water, one by A and four by B then each unit was necessary for the sufficiency of the set, so A 's unit was a cause of the harm occurring. This is a simple scenario like *Barnett* because only one sufficient set has actually occurred so there is no duplication. If A added 5 units and B simultaneously added 5 units then the scenario is equivalent to the double-hit hunters case of symmetrical duplication so both A and B would be labelled as causes on the NESS test because they are independently sufficient. If A added 5 units and B later added 5 units then A 's units may pre-empt B 's units from being causes so there is no duplication. Likewise if a total of seven units were added one after the other then the first five units may pre-empt the final two units from becoming causes. There is no danger of the 'at least enough' formulation being so expansive that it included cases of pre-emption because the NESS test correctly requires the unit to be necessary for a sufficient set of conditions that *actually occurred*. One ingredient in any sufficient set must be the condition that 'the damage has not already occurred'. It is impossible to construct a sufficient set containing some of A 's units and some of B 's in a pre-emptive situation like this, because as soon as A 's 5 units have caused the damage we lose the requisite 'the damage has not already occurred' condition in relation to any of B 's units.

The NESS test is very inclusive here, for example if one litre of pollution was sufficient to cause the injury and A added a teaspoon (five millilitres) of pollution while other parties added pollution up to a level of several litres, then the NESS test still leads us to the conclusion that

A 's teaspoon of pollution was a cause of the damage. Yet it is important to remember that this problem of asymmetrical duplication only arises when a) the harm is indivisible and b) the defendant contributes a source of a harmful agent that combines with other sources to form an indecipherable mass so that they operate simultaneously. It is also important to remember that causation is only one element of legal liability. Although liability would be joint and several, the defendant can seek contribution from the other responsible tortfeasors. Additionally, a defendant is not liable in negligence if his causal contribution was *de minimis*, so this reflects the idea that interpersonal responsibility is narrower than causal responsibility so a defendant will not be liable merely because he was a cause of the harm.

This section has addressed asymmetrical duplication which, along with symmetrical duplication, only arises in cases of indivisible damage. The next section looks at divisible damage to highlight the potential problems that can arise with the NESS test.

4.2.2 Causation of divisible damage

Divisible damage can be explained through a variation on the river pollution example. This time, instead of saying that an indivisible injury occurs at a level of five units of pollution, divisible disease is the equivalent of a situation where each unit of pollution will kill, for example, ten fish. If A adds one unit of pollution and B adds three units of pollution, then in total there will be four units of pollution so forty fish will die. The death of forty fish could be the outcome to which we apply the NESS test. In order for the set of factors to be sufficient to kill forty fish there must be four units of pollution so each unit was necessary for the sufficiency of the set. This would mean that A and B were both a cause of the outcome.

This does not seem to be the best way to understand the outcome. A corrective justice-based system of negligence liability is focused on interactions between individuals. Even though a claimant might be seeking compensation for his total loss, if this total loss consists of a number of smaller losses each of which is attributable to a different individual wrongdoer, then there

have been a number of individual interactions. An individual wrongdoer is only responsible for loss that he caused, and has no moral responsibility for entirely separate loss caused by a separate wrongdoer. So in the pollution example of divisible damage *A*'s one unit constituted a set that caused ten fish to die, and *B*'s three units constituted a separate set that was sufficient to cause thirty fish to die. The presence of *A*'s unit and *B*'s units are in no way related to each other, they have occurred independently, so their respective harms should also be treated independently. This is clearer in the following scenario: *A* shoots a walker and hits his arm, *B* shoots the walker and hits his leg. The shots were independent and the harms are independent. *A* caused the loss of the arm, *B* caused the loss of the leg. These examples seem potentially different because in the shooting scenario the causes remain separate and the effects are visibly separate too. In contrast, in the river scenario the causes cannot be separated once in the water and the effects are not visibly different. Although this means that *A*'s unit of pollution cannot be tied to ten specifically identified fish, this is a difficulty of proof and does not change the underlying causal model which is one where *A* has only caused the death of ten fish. This is important because in corrective justice-based interpersonal responsibility a person can only be morally responsible for an outcome if he was a cause of that outcome, so it would be wrong to impose moral responsibility on *A* for fish deaths that were caused by *B*.

In cases involving divisible harm the courts achieve corrective justice by ‘apportioning’ liability based on causal responsibility:

In tort law, the cause of action only arises upon the occurrence of “damage”: the legal forensic process is in fact best truly looked at in the reverse sequence...by defining at the outset the “damage” it is claimed has been caused by the fault of the defendant. In cases of divisible damage, legal responsibility for each part of the damage can then be traced back using the tortious mechanisms of causation, blameworthiness and duty.⁹³

⁹³ Simon Brown and Susan Lindsey, ‘Seeing the wood for the trees – Loftus-Brigham and apportionment of damage’ (2005) 21 Const LJ 431, 434.

But it is important to be clear that this is not an exception to joint and several liability, rather it is a recognition that each defendant caused a specific portion of the total loss so each is liable only for that loss that he caused. In *Holthby*,⁹⁴ the claimant developed asbestosis after having been exposed to asbestos over the course of his employment with different employers. He had worked for the defendant employer for approximately half of his working life. Asbestosis is a divisible disease which means that the defendant only caused a portion of the disease. Although they were not able to assess this contribution precisely, the court made a twenty-five percent reduction in damages, stating '[t]he court must do the best it can to achieve justice, not only to the claimant but the defendant, and among defendants'.⁹⁵ This is the correct approach to have taken, but the justification ought to have been made more forcefully. Apportionment achieves justice not simply because it seems unfair to make a single defendant shoulder the burden when other employers also caused the claimant's illness. Corrective justice demands that the defendant only pay for the loss that he has caused, and where the disease is divisible he has *only* caused a portion of the overall illness. As Hogg has argued, this means that a request from the defendant for apportionment should not be necessary but the 'apportionment' should take place as a matter of course.⁹⁶ He says that 'in many cases this will have to be a "best guess" approach, but to refuse to attempt such an apportionment is to saddle a sole defender with damages for losses which *he did not cause*'.⁹⁷

The precise contribution made by the defendant in *Holthby* could not be established, but justice was better achieved by taking a broad brush approach to attempt apportionment than if there had been no apportionment at all. This approach was followed in *Allen v British Rail Engineering Ltd (BREL)*,⁹⁸ where the claimant suffered 'vibration white finger' from the use of vibrating tools during his employment. The tools were used throughout his employment but this was only

⁹⁴ *Holthby* (n75).

⁹⁵ *ibid* [20] (Stuart-Smith LJ).

⁹⁶ Martin Hogg, 'Causation and apportionment of damages in cases of divisible injury' (2008) 12 Edin LR 99, 101.

⁹⁷ *ibid* 104 (emphasis added).

⁹⁸ [2001] EWCA Civ. 242, [2001] ICR 942.

negligent on the part of his employer once the risk of vibration white finger became known. The illness is one which worsens with the longer vibrating tools are used so it is divisible and apportionment was made even though the precise contribution of the negligence could not be established. Once again, corrective justice was more closely achieved by attempting imprecise apportionment than by denying apportionment completely.

4.2.2.1 Divisible damage and interaction between causes

The aetiology of divisible diseases is not always straightforward because there is not always a linear correlation between the potency of the cause and the severity of the disease. This is known as ‘effect modification’ but this is not an insurmountable challenge. Effect modification describes the phenomenon whereby a particular causal factor becomes more potent when combined with another causal factor and affects both divisible and indivisible diseases.⁹⁹ For example if smoking carries X relative risk of lung cancer, and exposure to asbestos carries Y relative risk of lung cancer, the relative risk of lung cancer in a smoker who has been exposed to asbestos is not necessarily simply X+Y, but may be higher because the interaction between smoking and asbestos may be an ‘effect modifier’. Similarly, in non-stochastic (divisible) diseases where the disease occurs along a continuum of severity, repeated exposure to one harmful agent, or exposure to that agent in combination with another harmful agent may increase the severity exponentially rather than incrementally. To return to the example of river pollution, say each unit of pollutant A would kill 5 fish, and each unit of pollutant B would kill 5 fish, but a combination of a unit each of pollutant A and B would kill 15 fish. The harm can still be divided into 5 dead fish for which A’s pollutant forms a NESS set, 5 dead fish for which B’s pollutant forms a NESS set, and 5 more dead fish for which the NESS set comprises both A and B’s pollutants.

Coggan and Taylor explain that the task of apportionment remains ‘relatively straightforward’ even in cases involving effect modification. This can be illustrated by considering the

⁹⁹ Coggan and Taylor (n70) 14.

apportionment that the court achieved in the case of *Rabman v Arearose Ltd.*¹⁰⁰ In this case the court undertook a detailed analysis of the various harms suffered by the claimant and determined which were causally attributable to the negligence of the first defendant alone, the second defendant alone and to both defendants' negligence together. The claimant was assaulted at work by two black youths who injured his right eye and it was held that the first defendant had negligently failed to protect him against this attack. The claimant then received negligent medical treatment from the second defendant which led to the permanent loss of sight in his right eye. It is clear that each defendant was causally responsible for distinct injuries, but the claimant also subsequently suffered various psychiatric illnesses and the court needed to apportion liability for these illnesses. The expert evidence showed that the claimant suffered a phobia of black people which was caused by the assault alone, a depressive disorder which was a reaction to the loss of the eye, post-traumatic stress disorder attributable to the assault and the operation, and enduring personality change which was due to the synergistic effect of the both causes. The court was thus able to apportion liability for each head of damages based on causal responsibility. It is clear that courts are able to undertake the task of apportionment in cases of divisible harm even where the various causes interact with a synergistic effect. This ought to be done in all cases of divisible harm in order to achieve corrective justice in these cases.

5. Using NESS to overcome common problems with exceptional legal tests

In negligence law the but-for test is supplemented with the *Wardlaw* test of 'material contribution to harm',¹⁰¹ so it is often said that the claimant must prove that the defendant's negligence 'caused or materially contributed to' the damage.¹⁰² It is unclear in the caselaw and academic literature whether the *Wardlaw* test is an application of the but-for test or an exception to it. If it is an exception to the but-for test it is further unclear whether it is an exception to the factual

¹⁰⁰ [2001] QB 351 (CA).

¹⁰¹ *Wardlaw* (n69).

¹⁰² ibid 620. Note that this phrase pre-dates the decision in *Wardlaw* but this is the leading case on the test of material contribution to harm.

causation requirement, or whether it is a test of factual causation that compensates for the conceptual inadequacy of the but-for test. This has led Honoré to criticise it as an ‘indefinite, if not indeterminate’ notion which is ‘purely pragmatic and leaves the theoretical problem untouched’.¹⁰³ This uncertainty, concerning just how exceptional the *Wardlaw* test is, has been important in the development of the law since the *Wardlaw* test was used as a kind of stepping stone between the but-for test and the *McGhee/Fairchild* test of ‘material contribution to the risk of harm’.¹⁰⁴ The uncertainty is most significant because it is unclear what causal problem is addressed by the test so its scope of application and its effects are ill-defined.¹⁰⁵

It will be argued in this section that the *Wardlaw* test is sometimes an application of the but-for test and sometimes an exception to it. Where it departs from the but-for test this is because of the conceptual inadequacies of the but-for test, so rather than being an exception to the causation requirement the *Wardlaw* test actually ensures that factual causation is tested more accurately than is allowed by the but-for test. This can be illustrated by drawing on the analysis of the NESS test that has been developed throughout this chapter and in section 4 in particular. Since the NESS test better reflects the underlying philosophical account of the concept of causation, it allows the lacunae in the but-for test to be pinpointed. In straightforward cases the but-for test works but only because it is shorthand for the NESS test, so it needs to be supplemented in more complex cases. It will be suggested that it is preferable simply to use the NESS test in those cases rather than to use the *Wardlaw* test which can only be understood properly by drawing on the NESS analysis and which risks adding to confusion by introducing its own vocabulary surrounding ‘contribution’.

¹⁰³ Honoré (n39) 364.

¹⁰⁴ The *McGhee/Fairchild* test is the subject of Chapter Four.

¹⁰⁵ It is worth noting that this is not identical to the American concept of a ‘substantial factor’ since that seems to apply only to cases of symmetrical duplication (see David W Robertson, ‘Causation in the Restatement (*Third*) of Torts: Three Arguable Mistakes (2009) 33 Wake Forest L Rev 1007, 1018). This means that the ‘substantial factor’ solution addresses the double-hit hunters problem of duplicative causation but it is not clear that it would apply in a case such as *Wardlaw* which involved a build up of the harmful substance in unknown quantities. The scope of the ‘substantial factor’ test is unclear because it has not been scrutinised in technical detail, instead ‘difficult puzzles...have been simply handed over to the jury’ (Stephen Bailey, ‘Causation in negligence: what is a material contribution?’ (2010) 30 LS 167, 169). Caution must therefore be exercised in referring to American scholarship on the practical implementation of the NESS test.

It is first necessary to present the facts and decision in *Wardlaw*, and to identify the specific areas of confusion surrounding its application and effects.

5.1 *Bonnington Castings v Wardlaw*

The claimant in this case was exposed to silica dust in the air at work coming from two sources. The majority of the dust, the ‘innocent’ dust, was produced by the hammer operated by the claimant and it was not possible to fit a dust extraction plant there. This was therefore a non-negligent source of silica dust in that it was unavoidable. The remainder, the ‘guilty’ dust, was present due to the negligence of the defendant in failing to provide adequate extraction at the swing grinders in the workshop. The claimant inhaled this silica dust in the air at work and developed pneumoconiosis, a disease of the lungs.

As Bailey has explained, the decision of the House of Lords cannot be fully understood without presenting the previous decisions and the point of appeal.¹⁰⁶ At first instance the Lord Ordinary had found that failure to provide respirators was a cause of the disease, which suggests that all the dust was considered ‘guilty’ at first instance. On appeal, the defendant argued that only the dust from the swing grinders was a result of negligence, and that the claimant had failed to establish causation in respect of this source of dust since the dust from the innocent source was far greater in quantity so was the ‘most probable’ cause of the disease. The Court of Session did not reach the question of whether the negligence concerned only the dust from the swing grinders, and disposed of the case on the point of causation. They held that exceptionally the onus of proof lay on the defendant to prove that the dust from the swing grinders could not have been a cause of the disease, and the defendant had failed to prove this. The issues to be decided in the House of Lords were therefore first whether the Court of Session had been correct to reverse the onus of proof, and second, if the onus of proof lay on the defendant, whether he could discharge this by proving that the innocent dust was the most probable cause because it was the far greater source of dust.

¹⁰⁶ Stephen Bailey, ‘Causation in negligence: what is a material contribution?’ (2010) 30 LS 167

The House of Lords held that the onus of proof was not reversed, it remained the task of the claimant to prove that the negligence had caused or materially contributed to the damage. On the facts, both sources had ‘materially contributed’ to the injury. The guilty dust had made a ‘material’, that is more than *de minimis*, contribution to the illness.

Lord Reid explained that the medical evidence was that pneumoconiosis is caused by ‘a gradual accumulation in the lungs of minute particles of silica inhaled over a period of years’.¹⁰⁷ He continued:

That means, I think, that the disease is caused by the whole of the noxious material inhaled and, if that material comes from two sources, it cannot be wholly attributed to material from one source or the other. I am in agreement with much of the Lord President's opinion in this case, but I cannot agree that the question is: which was the most probable source of the respondent's disease, the dust from the pneumatic hammers or the dust from the swing grinders? It appears to me that the source of his disease was the dust from both sources, and the real question is whether the dust from the swing grinders materially contributed to the disease. What is a material contribution must be a question of degree. A contribution which comes within the exception *de minimis non curat lex* is not material, but I think that any contribution which does not fall within that exception must be material. I do not see how there can be something too large to come within the *de minimis* principle but yet too small to be material.¹⁰⁸

Lord Keith noted also that the defendant's argument was based on the lack of evidence to show the precise proportions of dust coming from each source.¹⁰⁹

¹⁰⁷ *Wardlaw* (n69) 621 (Lord Reid).

¹⁰⁸ *ibid* 621 (Lord Reid).

¹⁰⁹ *ibid* 626,

5.1.1 The scope of the test

Part of the problem with this decision seems to be that the significance that it has acquired extends beyond the specific points of appeal that it actually addressed. The objective of the appeal was to determine whether the onus of proof was reversed, and whether the defendant could prove the absence of a causal link by showing that the innocent source was the ‘most probable’ cause. It was the reversing of the onus of proof by the Court of Session that was the unusual and unorthodox approach. By contrast, there was nothing controversial about holding that the claimant should prove that the negligence ‘caused or materially contributed to’ the disease. The claimant argued that ‘inhalation of the dust was the cause of the disease. If [the defendants] were shown negligently to have contributed a part of the dust which was not negligible, [the claimant’s] case was proved’.¹¹⁰ As Bailey has observed, ‘[n]o authorities were cited for this proposition and there was no suggestion that it was regarded as novel’.¹¹¹ This means that little attention was paid to the exact boundary between proof that the negligence ‘caused’ and proof that it ‘materially contributed to’ the disease, so the definitions and scope of application have been gleaned through subsequent analysis of the case and its application in later decisions. This means that factual aspects of the case, such as the ‘cumulative’ exposure to dust, and the divisibility of the disease, the lack of precise evidence as to the proportions of ‘guilty’ and ‘innocent’ dust, as well the absence of apportionment in the case, have been relied upon as determining the scope of the test of material contribution to harm.

5.1.1.1 Divisibility of damage

It is unclear what significance the divisibility of damage has for the applicability of the test of material contribution to harm. The court did not determine whether the claimant’s disease was divisible. It was called a ‘disease of gradual incidence’ which suggests it is divisible, but Lord Keith also said that without the negligent dust the claimant ‘would not have developed

¹¹⁰ ibid 617.

¹¹¹ Bailey (n106) 172.

pneumoconiosis when he did and might not have developed it at all’,¹¹² which suggests that it is indivisible.

Pneumoconiosis was later held to be divisible,¹¹³ so Hogg has suggested that *Wardlaw* ‘settled the point that, for a defendant to have caused a pursuer’s injury, it is not necessary that the defendant was the sole cause of the injury, but merely that, but-for the defendant’s conduct, the injury would not have occurred to the same extent’.¹¹⁴ If the fact that pneumoconiosis is divisible implies that the ‘material contribution to harm test’ only applies to divisible injuries then it was a mistake for the court not to have apportioned damages in *Wardlaw*.

Yet it has also been noted that where damage is divisible and the defendant has only caused a portion of the total loss there should be an apportionment of damages. The absence of apportionment in *Wardlaw* has led some to suggest that the disease was treated as though it were indivisible, so the divisibility of the disease is not the *reason* for the application of the test. Thus Smith LJ (extra-judicially) has said ‘material contribution to harm’ was a modification of the but-for test.¹¹⁵ In her view, the disease in *Wardlaw* was treated as indivisible but today it would be treated as divisible and apportioned. When the damages are apportioned it shows that the negligence was a but-for cause of a portion of loss. The absence of apportionment in *Wardlaw* could be taken as an indication that the but-for test was not satisfied in that case.¹¹⁶

Bailey notes, however, that damages had been assessed at first instance on the basis that provision of respirators would have prevented the claimant inhaling dust from both sources, so if damages were to be apportioned later the case would have to have been remitted to the Court of Session to consider evidence as to the appropriate measure of damages. It may have been the case that the state of the evidence would not have permitted such an assessment, and it

¹¹² *Wardlaw* (n69) 626.

¹¹³ *Nicholson v Atlas Steel Foundry & Engineering Co Ltd* [1957] 1 WLR 613.

¹¹⁴ Hogg (n96) 99.

¹¹⁵ Janet Smith, ‘Causation – the search for principle’ [2009] JPI Law 101, 101.

¹¹⁶ *ibid* 102.

‘presumably would not have been worth it in terms of cost’.¹¹⁷ This means that the case ‘cannot be read as holding that...the harm must as a matter of law be regarded as indivisible’.¹¹⁸ It has, however, been applied in cases of divisible and indivisible damage, for example the claimant in *Bailey* suffered an indivisible disease, brain damage.¹¹⁹ In this case, however, the causal process was still a ‘cumulative’ one in that the brain damage was a result of the claimant’s weakened state and this weakness had been caused by a number of factors. The next section therefore turns to consider the relevance of the causal process for the test of material contribution to harm.

5.1.1.2 ‘Cumulative’ causal process

In *Wardlaw*, Lord Reid observed that the disease was caused by ‘the whole of the noxious material inhaled’.¹²⁰ Lord Keith also noted that the negligent exposure was continuous over a long period of time and that ‘it was the atmosphere inhaled by the pursuer that caused his illness and it is impossible...to resolve the components of that atmosphere into particles caused by the fault of the defenders and particles not caused by the fault of the defenders, as if they were separate and independent factors in his illness’.¹²¹ This has been taken as suggesting that the test of material contribution to harm applies where the exposure to the harmful agent is ‘cumulative’ rather than ‘alternative’. These terms do not have an agreed meaning though.

Green suggests that the *Wardlaw* test applies to ‘cumulative causes’ but she equates ‘cumulative causes’ with ‘divisible injury’ so she takes the view that cumulative causes can be independent (such as in *Performance Cars*¹²²) or interdependent as in *Bailey*.¹²³ The claimant in *Bailey* was in a weakened state which, when she vomited, caused her to be unable to respond naturally to the vomit, instead aspirating it leading to cardiac arrest and brain damage. The defendant hospital

¹¹⁷ *Bailey* (n106) 173.

¹¹⁸ *ibid.*

¹¹⁹ *Bailey v Ministry of Defence and another* [2008] EWCA Civ. 883, [2009] 1 WLR 1052

¹²⁰ *Wardlaw* (n69) 621.

¹²¹ *ibid* 626.

¹²² See Chapter One (n137)

¹²³ Sarah Green, ‘Contributing to the risk of confusion? Causation in the Court of Appeal’ (2009) 125 LQR 44.

had negligently failed to resuscitate her adequately following an operation, thus causing her to be weaker than she would otherwise have been. At the same time she developed pancreatitis which also caused her to become weaker but this was a natural complication and not attributable to the negligence of the hospital. The question was whether the negligent failure to resuscitate the patient after her first operation was a cause of her ultimate injury. Green argues that *Bailey* effectively involved divisible harm because ‘weakness was undoubtedly divisible, since patients are not classed simply as being either weak or strong, but lie somewhere along an almost infinitely calibrated spectrum’.¹²⁴ However, in *Bailey*, the weakness was the cause of the harm which was the claimant’s eventual brain damage due to aspiration of her vomit. The damage here is indivisible, even though its cause – weakness – came from two sources with each contributing to the degree of weakness. Others regard the causal process as forming the distinction between divisible and indivisible injury, for example Gullifer has suggested that harm is divisible if there is a temporal or spatial separation between the causes as in *Performance Cars*, and indivisible if the candidate causes are cumulative causes of the same type that combine together as in *Wardlaw*.¹²⁵ Stauch says that where the disease is indivisible it should not matter whether the causal process is cumulative or alternative.¹²⁶

It is impossible to determine the scope of the test from the facts of *Wardlaw* alone because the limited grounds of appeal mean that the above issues simply were not addressed.

5.1.2 Analysis of the *Wardlaw* test

Bailey seeks to show that the test of material contribution to harm is an application of the but-for test. It is important that he regards the but-for test as being synonymous with factual causation, so any departure would be exceptional. For him, the but-for test must be satisfied, and when we say that a claimant must prove that the negligence ‘caused or materially contributed to’

¹²⁴ ibid 45.

¹²⁵ Louise Gullifer, ‘One cause after another’ (2001) 117 LQR 403, 406.

¹²⁶ Marc Stauch, “‘Material contribution’ as a response to causal uncertainty: time for a rethink” (2009) 68 CLJ 27, 29.

the loss, this is a more elaborate way of saying that the claimant must prove but-for causation that resolves ‘the need to find words that cover both cases arising (effectively) from single causes and those arising from multiple causes’.¹²⁷ We know, however, that an outcome does not have a single cause. This means that when Bailey says that ‘caused’ relates to cases arising *effectively* from single causes he is making recourse to the Hart and Honoré type approach that uses common sense to distinguish ‘causes’ from ‘mere conditions’. He is selecting between all of the necessary conditions, so his approach is evaluative rather than purely factual. This is why he finds it significant to remind us that in any given causal set, more than one factor may be identified as ‘a cause’ rather than as a ‘mere condition’.

The NESS analysis in sections four and five showed that there are a range of causal problems that the but-for test is unable to resolve because of its conceptual inadequacies. Causation, as with any aspect of negligence, can also give rise to evidential problems, and in *Wardlaw* Lord Keith noted that the defendant’s case rested partly on the fact that it was not possible to prove the precise proportions of silica dust attributable to each source. The defendant had also argued that the greater source of dust was the ‘most probable’ cause, but this confuses the question of what the claimant must prove with the standard to which he must prove it i.e. on the balance of probabilities.¹²⁸ It is therefore also possible that the test of material contribution to harm seeks to reinforce this distinction and reiterate that an outcome is caused by a set of conditions and each condition is equally deserving of the label ‘cause’ as Mill showed in his work on the philosophy of causation.

The meaning of the *Wardlaw* test of ‘material contribution to harm’ will necessarily be obscured if it responds both to conceptual inadequacies of the but-for test and to evidential difficulties. The preferable solution is therefore to adopt the NESS test since it is conceptually more sound than the but-for test and able to resolve causal problems that the but-for test cannot. This would mean that evidential issues arising from the balance of probabilities standard of proof, or from

¹²⁷ Bailey (n106) 178.

¹²⁸ See also Bailey (n119) [56] (Waller LJ).

weaknesses in the evidence available to the court, could be readily identified for what they are. The final section of this chapter presents a practical example of a recent decision where the *Wardlaw* test was applied and seeks to illustrate the advantages of using the NESS test.

5.2 Practical application of the NESS test

The NESS analysis of causation enables the causal question to be clearly articulated, making the resolution of cases more straightforward. This final section considers a number of cases where there has been confusion in application of the *Wardlaw* test and in the question of ‘apportionment’ and seeks to show that approaching these cases via the NESS test clarifies the issues.

5.2.1 The question of causation in *Bailey*

The decision of the Court of Appeal in *Bailey* was notable for showing confusion over the meaning and applicability of the *Wardlaw* test of material contribution to harm. Waller LJ, with whom Sedley and Smith LJJ concurred, approached the issue by asking ‘was this a case in which the judge was entitled to depart from the but-for test?’¹²⁹ As this chapter has argued, the legal issues cannot be fully articulated if the but-for test is taken as the frame of reference, because the but-for test is conceptually inadequate. Using the NESS test makes the resolution of the case simpler and reduces the potential for future confusion.

As noted above, the claimant in *Bailey* was in a weakened state and therefore aspirated her vomit leading to cardiac arrest and brain damage. The defendant’s negligence had led to weakness, and she was also weakened by a naturally occurring illness. The question was whether the negligent failure to resuscitate the patient after her first operation was a cause of her ultimate injury.

The first step is to determine whether the damage is divisible; the damage in *Bailey* was the brain injury which in this case was indivisible. The cause of the brain damage was the claimant’s

¹²⁹ *Bailey* (n119) [36] (Waller LJ).

weakened state which prevented her from responding naturally to her vomit. Applying the NESS test, the question is whether the weakness that was due to the defendant's negligence was a necessary element of a sufficient set that *actually occurred*? Since the patient gradually became weakened and this overall weakness led to her failing to respond to her vomit, it is clear that the weakness formed an 'indecipherable mass' because the effects operated on her simultaneously. It was the overall weakened state that caused the damage, and the negligence contributed to the total level of weakness, so it was a cause of the claimant's brain damage. The judge at first instance seems to have understood that this was a relatively simple issue, explaining that although he could not say 'whether the contribution made by [the negligence] was more or less than that made by the pancreatitis...the natural inference is that each contributed materially to the overall weakness and it was the overall weakness that caused the aspiration', so he found that the causal link had been established.¹³⁰

In the Court of Appeal, Waller LJ eventually concluded that causation could be established using the test of material contribution to harm, but his reasoning was more complex since he considered this test to be an exception to the but-for test. He therefore asked whether it was a case in which he was entitled to depart from the but-for test which led him to weigh up the policy arguments and to seek to distinguish the case from a range of other exceptional approaches such as the loss of chance argument and the *Fairchild* exception. Waller LJ saw the following issue as central:

It is important to be clear precisely what *Wardlaw* decided. Did it decide that in a cumulative cause case where the inadequacies of medical science meant that the relative potency could not be established all a claimant had to establish was a 'material' contribution?...or did a claimant still have to establish that 'but for' the contribution of the negligent cause, the injury would not have occurred?¹³¹

¹³⁰ *Bailey v Ministry of Defences, Portsmouth Hospitals NHS Trust* [2007] EWHC 2913 (QB), [61] (Foskett J).

¹³¹ *Bailey* (n119) [39] (Waller LJ)

This phrasing of the issue has a number of weaknesses: it assumes that proof of a material contribution does not satisfy the but-for test; it does not address the nature of the damage; and it assumes that the *Wardlaw* approach was necessitated because it involved cumulative causes and because the relative contributions could not be measured with precision. Waller LJ did not engage with these questions in any depth, but was satisfied that Lord Rodger had established in *Fairchild* that ‘in cumulative cause cases such as *Wardlaw* the but-for test is modified’.¹³² As discussed in this chapter, however, the NESS analysis shows that where damage is divisible ‘material contribution to harm’ translates as applying the but-for test (as shorthand for the NESS test) to a portion of the overall damage, and that where the damage is indivisible it also translates as applying the but-for test where the victim was only exposed to the threshold amount of the harmful agent. It is only where indivisible damage occurred and the causes are duplicative that a test of material contribution would be an exception to the but-for test. But in these cases, if the *Wardlaw* test is applied consistently with the NESS test, then it makes up for the conceptual inadequacy of the but-for test and is not an exception to the factual causation requirement. This means that there is no need to consider, as Waller LJ did, whether a departure from the but-for test is justified on policy grounds. Waller LJ considered arguments that have been raised in the medical negligence context, but that relate to claims based on the loss of chance idea or on the *Fairchild* exception, and these arguments *are* exceptional. In other words, by failing to observe that the but-for test is conceptually inadequate and needs to be supplemented in order to accurately test factual causation, Waller LJ made life difficult for himself and weighed up policy factors that extend beyond corrective justice and which are unnecessary.

5.2.2 The question of ‘apportionment’

Another problem arising out of the uncertainty surrounding the reasons for the application of the *Wardlaw* test is that it has led to uncertainty whether ‘apportionment’ of damages is

¹³² Ibid [43] referring to *Fairchild v Glenhaven Funeral Services* [2002] UKHL 22, [2003] 1 AC 32, [129] (Lord Rodger)

appropriate. If the courts apply the test of material contribution because of a lack of evidence as to the precise proportions of the harmful agent attributable to each source then, it has been argued, it would be inappropriate for the court to apportion liability because it would lack a rational basis on which to calculate the portion of the damage that is attributable to the defendant's negligence. As already discussed, damages should be 'apportioned' where the injury is divisible. Since each source of the harmful agent causes a portion of the total loss, this is a false apportionment – it is not an apportionment of liability but a question of quantifying that portion of the loss attributable to the defendant's negligence. It is therefore the definition of the damage, rather than the NESS test, that determines the 'apportionment' issue. The NESS test, however, is still valuable because it highlights that the *Wardlaw* test is needed to respond to a conceptual weakness in the but-for test and not because of an evidential problem.

In *Hatton v Sutherland*,¹³³ a case involving psychiatric illness resulting from occupational stress, the claimant established causation by showing that the defendant's negligence made a material contribution to her illness and in the Court of Appeal Hale LJ concluded that where there were multiple causes of a psychiatric injury then apportionment should be attempted. In the decision of the Court of Appeal in *Dickins v O2 plc*,¹³⁴ however, it was suggested that where the test of 'material contribution to harm' is applied it would be inappropriate to apportion liability regardless of whether the harm was divisible or indivisible. The reason for this conclusion was the court's understanding that the 'material contribution to harm' test is used where the various contributions made by the defendant(s) cannot be quantified.

[I]n a case which has had to be decided on the basis that the tort has made a material contribution to harm but it is not scientifically possible to say how much that contribution is (apart from the assessment that it was more than *de minimis*) and where

¹³³ [2002] EWCA Civ 76, [2002] 2 All ER 1.

¹³⁴ [2008] EWCA Civ 1144, [2009] IRLR 58.

the injury to which it has lead is indivisible, it will be inappropriate simply to apportion the damages across the board.¹³⁵

Since the extent of the contribution cannot be quantified, the court reasoned that any apportionment would necessarily be artificial:

That means that a claimant can succeed on causation even though he cannot demonstrate what the causative potency of the tort was, save to say that it had some effect beyond the minimal. It seems to me that, if in one breath the judge holds that all that can be said about the effect of the tort is that it made an unspecified material contribution, it is illogical for him, in the next breath, to attempt to assess the percentage effect of the tort as a basis for apportionment of the whole of the damages.¹³⁶

It is not at all clear that the reason for applying the *Wardlaw* test is that the relative contributions cannot be quantified. As this chapter has illustrated, the *Wardlaw* test may also represent an application of the but-for test to a portion of the total loss in cases of divisible damage, and may make up for the conceptual inadequacies of the but-for test in over-determined cases of indivisible damage. Since the nature of the problem addressed by the *Wardlaw* test is unclear it is natural that there will be confusion in its application. As this chapter has shown, the first step in any case is to define the damage as being divisible or indivisible. Once this has been done, the NESS test can be applied to either type of damage and can resolve problems of pre-emption and duplication. If the damage is indivisible then the negligence is a cause of the whole of the loss so there is no apportionment vis-à-vis the claimant, and joint and several liability applies. If the damage is divisible and the negligence has caused only a portion of the total loss then apportionment is required in principle because the defendant is only causally responsible for a portion of the loss so only that portion of the loss is ‘wrongful’ in corrective justice. If there is an insurmountable evidential barrier to the apportionment exercise then this must be addressed squarely.

¹³⁵ ibid [46] (Smith LJ).

¹³⁶ ibid [43] (Smith LJ).

Chapter One showed that causation has a vital but limited role in interpersonal responsibility and this chapter has drawn this issue out in greater detail to isolate factual causation, the fact of being *a cause*, from evaluative conclusions that a condition was *the responsible cause*. It was shown that the NESS test is preferable to the but-for test as a test of factual causation because it is better matched to the philosophical account of what it means to be a cause and is therefore able to resolve more complex causal problems that the but-for test is unable to resolve. Adopting the NESS test would ensure that courts are equipped to address a wider range of causal scenarios without resorting to exceptional tests. This is important because it has shown that cases such as *Wardlaw* and *Bailey* may involve an exception to the but-for test but they do not involve an exception to the causation requirement. Having addressed the conceptual aspects of causation in this chapter the remaining chapters turn to evidential problems that have arisen and consider how they can be resolved consistently with principles of corrective justice. The work in this chapter will be valuable because the NESS analysis means that the evidential problems can be more clearly identified. The final chapters will also be strengthened by being able to draw on the analysis of the role that causation has within interpersonal responsibility and of the roles of the other negligence doctrines in addressing the remaining aspects of interpersonal responsibility.

Chapter Three: Loss of a chance

Introduction

The loss of a chance argument has been raised by claimants who have been unable to prove on the balance of probabilities that the defendant's negligence was a cause of their physical damage. The claims in *Hotson v East Berkshire Health Authority*,¹ and *Gregg v Scott*,² arose in the medical negligence context and in both cases the negligence consisted of the doctor's failure to diagnose and treat an existing illness. In both cases the court found that on the balance of probabilities the patient's existing illness could not have been cured with careful treatment, so the delay was not a cause of any eventual physical damage. The claimants argued instead that although it was less than probable that their condition could have been cured, the 'chance' of a cure had been reduced by the negligence and this 'chance' was something of value. They therefore sought to reformulate the actionable damage as the 'chance' of avoiding physical injury rather than the physical harm itself. If it succeeded, this action for loss of a chance would lead to proportionate recovery of damages for the physical harm, for example in *Hotson* the loss of a 25 percent chance of avoiding avascular necrosis was valued at 25 percent of the total loss caused by the avascular necrosis. Although this argument therefore centres on the question of whether a 'chance' of a better outcome can constitute actionable damage in negligence, it is an attempt to sidestep the difficulties of proof of causation of the physical harm on the balance of probabilities standard of proof.

It will be argued in this chapter that the House of Lords was right to reject the claim for proportionate recovery for the loss of a chance of a better medical outcome in both of these cases. In corrective justice-based interpersonal responsibility, liability is for 'wrongful loss' and the loss can only be characterised as wrongful if it was, in fact, caused by the defendant's

¹ *Hotson v Fitzgerald and others* [1985] 1 WLR 1036 (QBD); *Hotson v East Berkshire Health Authority* [1987] 2 WLR 287 (CA); *Hotson v East Berkshire Health Authority* [1987] AC 750 (HL).

² *Gregg v Scott* [2002] EWCA Civ 1471, [2003] Lloyd's Rep. Med. 105 (CA); [2005] UKHL 2, [2005] 2 AC 176 (HL).

negligence. Causation is a factual relationship, so the negligence either was or was not a cause of the harm. The claimant need only persuade the court of this fact on the balance of probabilities and if the court is, on balance, persuaded that the negligence was a cause of the damage then it accepts the fact of causation to be proved and the claimant recovers in full. Although the loss of chance argument ostensibly reformulates the damage as the chance of avoiding physical harm, it will be argued that what it in fact achieves is to ‘discount’ liability to reflect the degree of doubt over the fact of causation.

However, it will also be argued in section 4.2 that the doctor/patient relationship raises unique issues of interpersonal responsibility because the patient’s ‘chance’ is valued by both parties and has intrinsic value independently of the physical outcome. At the time of diagnosis and treatment, the outcome of treatment is uncertain and beyond the control of the parties – the doctor and patient both regard him as having a ‘chance’ in these circumstances and they do what they can to improve his chance – the doctor gives him careful treatment, indeed the purpose of his duty of care when a patient is already ill is to take an uncertain situation and do what he can but he cannot guarantee cure. Both parties therefore take the statistical medical assessment of the likelihood of cure as being something that is valuable to the claimant so a reduction in the statistical chance of cure is a loss of something of value and ought to be recognised as actionable damage. Importantly it is valued for its own intrinsic worth rather than as a proportion of the physical harm. While the practical effect is that loss of chance would ‘assist’ claimants by providing compensation even where the doctor’s negligence is not proved to have caused physical harm, this compensation would be very modest and is certainly not a way to bypass the causation requirement regarding the physical harm.

1. Analysis of *Hotson v East Berkshire Health Authority*

The purpose of the following section is to set out the facts of *Hotson* and to clarify some of the issues that arose in that case. The claim for the loss of a chance of avoiding injury was successful

at first instance and in the Court of Appeal but was ultimately unsuccessful in the House of Lords. A later section of the chapter will seek to evaluate the loss of a chance argument. The current section takes the preliminary step of setting the facts of the case within the framework of the conventional understanding of negligence. The purpose is to set aside any possible misconceptions about how the traditional principles of causation operate. This includes building on Chapter One to separate clearly the concepts of damage, causation and quantification. This will allow an analysis of the loss of chance argument to be based on a clear understanding of the concepts involved and of precisely where the problem lies. Three main strands of the case emerge here: the definition of damage and its relation to the so-called ‘hook’ argument; the relationship between damage and quantification, particularly in relation to the vicissitudes principle; and the different role played by damage in contract and tort. The relationship between a statistical assessment of the probability of causation and the balance of probabilities standard of proof will be considered in a later section in relation to both *Hotson* and *Gregg*.

1.1 The facts

The claimant in *Hotson* was a boy who had fallen from a tree and sustained an injury to his hip, specifically a fracture to the left femoral epiphysis. He was taken to hospital (for which the defendant health authority was responsible) where he was examined, his knee was x-rayed and bandaged. The medical staff negligently failed to x-ray his hip so they did not diagnose the injury to his hip and instead sent him home untreated. He continued to suffer severe pain for the next five days after which he returned to hospital where his hip was x-rayed, the correct diagnosis was made, and emergency treatment begun. The boy suffered avascular necrosis in the epiphysis which involved deformity of the hip, with loss of movement and a limp, and would worsen in future due to osteoarthritis developing in the joint because of the injury. The femoral epiphysis is a layer of cartilage between the head and neck of the femur and avascular necrosis of the epiphysis develops because of a failure of the blood supply to the epiphysis. The award of

damages for the five day period of suffering was unproblematic. The difficulty arose in relation to the avascular necrosis. The injury clearly constituted actionable damage, the defendant owed the claimant a duty of care and had breached that duty. The difficulty was in establishing on the balance of probabilities that the negligence was a cause of the injury. At first instance, Simon Brown J found that there was only a 25 percent chance that the claimant would have avoided avascular necrosis if his injury had been treated appropriately, so a conventional claim would have failed. However, he awarded the claimant damages to compensate the loss of a 25 percent chance of avoiding avascular necrosis. These damages were 25 percent of the value of the overall injury. He effectively treated the case as involving an issue of quantification of the loss caused by the delay in diagnosis, leading to proportional recovery, rather than as involving a difficulty of establishing a causal link between the delayed diagnosis and the final injury. The Court of Appeal upheld the claim, but it was rejected by the House of Lords. The House of Lords did not reject the concept of the loss of a chance, but held that in this instance the finding of fact was that, on the balance of probabilities, at the time of the negligence insufficient blood vessels remained intact for the injury to be treatable. Thus he was destined to suffer avascular necrosis and there was no ‘chance’ that had been lost.

With regard to the causal process involved in this case, the judge at first instance had found that avascular necrosis occurs when there are insufficient blood vessels to keep the epiphysis alive. When the claimant first went to hospital some blood vessels remained intact although distorted, but the delay in commencing treatment meant that already ruptured blood vessels continued to bleed into the joint thus increasing pressure in the joint and blocking the intact blood vessels. The delay in diagnosing and treating the injury therefore caused damage to those blood vessels that had remained and sealed the claimant’s fate by turning his injury into an inevitability.³ The judge thought it ‘possible but improbable’ that sufficient vessels remained intact after the fall because the injury to the joint was at the upper end of the spectrum of severity. He found that

³ *Hotson v Fitzgerald and others* [1985] 1 WLR 1036 (QBD), 1040 (Simon Brown J).

there was a 25 percent chance that following the fall sufficient blood vessels had remained intact to keep the epiphysis alive and avoid avascular necrosis. Lord Mackay, in the House of Lords, suggested that Simon Brown J had arrived at this figure because it was midway between the figures given by the two expert witnesses.⁴ Simon Brown J said himself that he was ‘unattracted to, and finally unable to accept, either of the competing extreme views’,⁵ so his conclusion does seem to achieve a compromise between the views of the two expert witnesses.

1.2 Distinguishing damage, causation, and quantification

As argued in Chapter One, it is essential to maintain a clear distinction between the negligence doctrines, notably damage, causation, and quantification. At first instance, Simon Brown J said that ‘[i]n the end the problem comes down to one of classification. Is this on true analysis a case where the plaintiff is concerned to establish causative negligence or is it rather a case where the real question is the proper quantum of damage? Clearly the case hovers near the border’.⁶ The conventional approach to negligence is that a claimant must show that he has suffered actionable damage and that the negligence was a cause of this damage. Physical injury is a recognised form of damage, but the loss of a chance of avoiding physical injury was previously not recognised as being capable of forming the gist of a negligence action. Once all the elements of a negligence claim have been established and the court is concerned with valuing the claimant’s loss, the principle of *restitutio in integrum* means that the claimant should be compensated for all the losses that flow from the personal injury. Certain lost chances are recoverable at this stage. Simon Brown J explained:

Time and time again courts evaluate past and future medical risks and award damages based on an assessment of the likelihood (a) of some adverse medical condition, like epilepsy or osteo-arthritis, developing consequent on the injury, or (b) that some pre-

⁴ *Hotson v East Berkshire HA* [1987] AC 750 (HL).

⁵ *Hotson v Fitzgerald and others* [1985] 3 All ER 167, 173 (Simon Brown J’s evaluation of the evidence is available in this report but omitted from that of the Weekly Law Reports).

⁶ *Hotson* (QBD) (n3) 1043.

existing, perhaps degenerative, condition would in any event have manifested itself so as to cause the same or at any rate some lesser degree of disability as has been occasioned by the injury.⁷

This is simply an aspect of returning the claimant to his pre-tort position. But the judge then went on to overstate the significance of this, suggesting that:

There is really no significant difference between that exercise and what the court is being invited by the plaintiff to do in the instant case.⁸

In fact, there is a significant difference between these principles and the loss of a chance argument in that they address issues of quantification whereas the loss of chance argument involves a prior issue of damage and causation. The following sections will therefore distinguish the aspects of damage, causation and quantification in those traditional principles in order to show that the loss of a chance of avoiding avascular necrosis is not analogous with existing negligence principles.

1.2.1 The ‘hook’ argument

Simon Brown J considered that the claimant’s five days pain and suffering constituted actionable damage, and that the lost chance of recovery could simply be regarded as a question of quantification.⁹ This is referred to by Stapleton as the ‘hook’ argument, the idea being that once it has been proved that the negligence caused some physical damage, albeit minimal, any lost chance of avoiding further harm can be ‘hooked’ on to this claim as an issue of quantification.¹⁰ But he went on to say ‘this very point underlines how unsatisfactory it would be to suppose that the case should turn entirely on whether there is any directly provable injury, however slight.

⁷ *ibid* 1045.

⁸ *ibid.*

⁹ *Hotson* (QBD) (n3) 1045.

¹⁰ Jane Stapleton, ‘Cause in fact and the scope of liability for consequences’ (2003) 119 LQR 388; ‘Loss of the chance of cure from cancer’ (2005) 68 MLR 996.

That will itself often be a matter of chance'.¹¹ However Simon Brown J has overstated the 'hook' concept here. In his earlier explanation of conventional principles he had noted that courts regularly assess the likelihood that the claimant will develop a medical condition *consequent* on the injury.¹² It is necessary to show that there is a causal link between the injury that the defendant has actually caused and the anticipated future illness. The defendant is not required to pay for all future illnesses that the claimant might suffer, he is required to compensate for those illnesses that the claimant might suffer *because* of the harm the defendant has negligently caused him. In *Holson* the claimant was unable to prove that the delay in treatment had caused the avascular necrosis, so he would have been equally unable to prove that his five days pain and suffering was causally related to the avascular necrosis. Lord Bridge explained this in the House of Lords:

The damages referable to the plaintiff's pain during the five days by which treatment was delayed in consequence of failure to diagnose the injury correctly, although sufficient to establish the authority's liability for the tort of negligence, have no relevance to their liability in respect of the avascular necrosis. There was no causal connection between the plaintiff's physical pain and the development of the necrosis. If the injury had been painless, the plaintiff would have to establish the necessary causal link between the necrosis and the authority's breach of duty in order to succeed. It makes no difference that the five days' pain gave him a cause of action in respect of an unrelated element of damage.¹³

By understanding that the hook argument would only succeed where there is a causal relationship between the actionable injury and the risk of future illness it is clear that it is not arbitrary as Simon Brown J suggested. The hook idea should continue to be used where actionable injury causes the claimant an increased risk of future illness, but cannot apply in

¹¹ *Holson* (QBD) (n3) 1045-46.

¹² See text to (n7) (emphasis added).

¹³ *Holson* (HL) (n4) 780 (Lord Bridge).

Hotson where there is no causal relationship between the five days pain and suffering and the increased risk of avascular necrosis.

1.2.2 The vicissitudes principle

As noted above, Simon Brown J also observed that courts regularly take into account the likelihood that some pre-existing condition would have resulted in the same degree of disability when valuing the loss caused by the defendant.¹⁴ This is known as the vicissitudes principle. When calculating loss of earnings the multiplier used by the court is reduced to take account of the ‘vicissitudes of life’ i.e. the risk that the claimant’s earnings would have been reduced by other unrelated accidents or illnesses.¹⁵ Simon Brown J said that this amounts to holding that ‘if the risk, or chance [of avoiding injury], is less than 50 per cent then the plaintiff gets nothing. If, however, it is over 50 per cent then the court should proceed to determine the matter as one of quantum, which involves having regard to the chances and contingencies and making discount accordingly’.¹⁶ This, he said, ‘smacks somewhat of heads I win, tails you lose’.¹⁷

He is right to say that such a position would discriminate in favour of defendants, yet once again he has overstated the issue. The discount made for the vicissitudes of life is made to reflect as accurately as possible the true value of the loss suffered by the claimant. Damages should return the claimant to the position he would have been in without the negligence so the court must do its best to account for the events that would have befallen him anyway including the likelihood that unrelated illness or accident would have impacted on the claimant’s earning capacity in the future. It is an issue of quantification. The fact that the defendant’s negligence was a cause of the injury actually suffered is not in doubt. In *Hotson* the claimant did not establish that the defendant’s negligence caused him to suffer avascular necrosis so the subsidiary task, of taking

¹⁴ See text to (n7)

¹⁵ *Lim Poh Choo v Camden & Islington AHA* [1980] AC 174 (HL)

¹⁶ *Hotson* (QBD) (n3) 1049.

¹⁷ *ibid* 1049.

the pre-existing chance of injury into account when valuing the loss caused by the negligence, never arises.

The difference is apparent if *Hotson* is compared with *Smith v Leech Brain*.¹⁸ In that case the defendant's negligence caused the claimant to suffer a burn on his lip which triggered a pre-cancerous condition and led to his eventual death. The pre-cancerous condition meant that independently of the negligence there was an increased likelihood that he would have developed cancer in the future, so his damages were reduced by five percent to account for this possibility. Stapleton suggests that this is indistinguishable in effect from the loss of chance-based claim for proportionate recovery:

The “discount” is made not to reflect that chance that the triggering had been due to a cause other than the defendant’s fault (because it was clear that this had not been the case) but to reflect the true value of the loss which the defendant had caused the plaintiff to suffer. In many cases, if not all, this approach will give a result indistinguishable from that produced had the claim been framed in terms of loss of a chance.¹⁹

Even though the *result* might look indistinguishable from a claim for proportionate recovery framed as a claim for the loss of a chance, it does not mean, as Stapleton suggests, that valuing a lost chance as a proportion of the physical harm to which the chance relates is conceptually sound. There was no doubt in *Smith* that the defendant’s negligence had caused the claimant to suffer the harm at that particular time. Yet when making a discount for the vicissitudes of life at the valuation stage, account was taken of the pre-cancerous condition to personalise the likelihood that the claimant would have died prematurely anyway. In *Hotson* the doubt surrounded the question of whether the defendant’s negligence had caused him to suffer avascular necrosis at this time. If proportional recovery were allowed in *Hotson* it would not reflect the chance that the claimant would have suffered the same loss due to an unrelated illness

¹⁸ *Smith v Leech Brain & Co* [1962] 2 QB 405 (QBD).

¹⁹ Jane Stapleton, ‘The gist of negligence: Part 2 the relationship between “damage” and causation’ (1988) 104 LQR 389, 398.

at some stage in the future, it would reflect the degree of doubt over whether the defendant had even caused him any loss.

In the House of Lords, Lord Ackner correctly stated that ‘[o]nce liability is established, on the balance of probabilities, the loss which the plaintiff has sustained is payable in full. It is not discounted by reducing his claim by the extent to which he has failed to prove his case with 100 per cent certainty’.²⁰ Understanding that the vicissitudes principle relates to valuation of loss once a causal link has been established makes it clear that it is distinct from the loss of chance argument which seeks to solve a difficulty relating to proof of causation. Simon Brown J suggested that the rejection of the loss of chance argument would be inconsistent with the conventional approach in this respect, yet it has been shown that this is not the case.

Stapleton has criticised the decision of the House of Lords for failing to resolve the essence of the claimant’s argument, ‘namely whether reformulation of the gist in terms of loss of a chance should now be acceptable’.²¹ It is clear that this is what the claimant was trying to do, and that he was trying to do this because of his inability to prove a causal link between the negligence and the physical outcome. The above discussion should also have added some clarity to what happens at each stage of a conventional claim in terms of damage, causation and valuation, so that arguments for and against reformulating the gist in terms of loss of a chance are considered against an accurate picture of traditional recovery.

1.3 The contract/tort distinction

At first instance Simon Brown J noted that loss of chance is actionable in contract law but rejected the contract/tort distinction as a basis on which to reject loss of chance. In his view, the particular branch of the law of obligations in which the claim is brought is not relevant to weighing the merits of the case since ‘[t]his would be entirely artificial, as is well illustrated by

²⁰ *Hotson* (HL) (n4) 793 (Lord Ackner).

²¹ Stapleton, ‘The gist of negligence’ (n19) 393.

recognising that it would turn, for instance, on whether the patient was consulting his doctor privately or under the National Health Service.²²

In principle a coherent distinction can be drawn between the position in contract and in tort because the doctrine of damage has a different role in each of these branches of law. For a claim to succeed in the tort of negligence the claimant must bring together the elements of actionable damage, duty of care, breach of duty, factual and legal causation. This means that he must prove that he has suffered a legally recognised form of loss in order to have a complete claim. Only once he establishes all the elements of the tort does the question of valuing the claim arise. In contrast, in contract law a claim becomes actionable upon breach of contract which means that loss suffered by the claimant is relevant only to quantification and is not a necessary element for him to have a successful claim in the first place.

Simon Brown J's rejection of the contract/tort distinction is specific to the medical negligence context since the same doctor/patient relationship can arise in a contractual or non-contractual relationship. In this particular context, allowing recovery for loss of a chance in contract but not in tort would lead to disparity in the treatment of claims depending on whether the patient was treated by the National Health Service or a private healthcare provider. While it may be the case that a patient can choose to pay for private healthcare and thereby purchase a higher quality of care, it would seem nonsensical to suggest that the actual losses that a patient suffers are different depending on whether they had a contract or not. As Reece has argued, '[t]he private patient may have paid for a higher standard of care, but the standard of proof should surely remain the same'.²³

The real question is whether the 'chance' that a patient has of recovering from illness or injury when he visits his doctor is something that is capable of constituting a loss. The starting point for analysis must be the concept of chance rather than the type of legal action. So whilst it is conceptually correct to allow for a different approach in tort and in contract because of the

²² *Holson* (QBD) (n3) 1044.

²³ Helen Reece, 'Losses of Chances in the Law' (1996) 59 MLR 188, 191.

different role played by damage within the different types of claim, the important question is whether the ‘chance’ of recovering from illness is a meaningful concept, something of which the patient can be deprived and in respect of which he ought to be able to claim against his doctor.

2. Analysis of *Gregg v Scott*

The loss of chance argument for proportionate recovery was subsequently raised again in the medical negligence context in the case of *Gregg v Scott*.²⁴ The claim once again failed. The claimant in *Gregg* went to the defendant doctor complaining of a lump under his arm which the defendant negligently diagnosed as a benign fatty lump rather than referring the patient for further tests which would have shown that it was cancerous. This resulted in a nine month delay in the diagnosis and commencement of treatment of the claimant’s disease. The claimant argued that the nine month delay due to the defendant’s negligence had caused him to lose a chance of being cured because his chance of cure had been reduced from a statistical 42 percent at the time of negligence to 25 percent at the time of trial.²⁵ The medical definition of cure is ten years disease-free survival, so the claim was that the negligence had caused him to lose a chance of ten years disease-free survival.²⁶ (Since the details of the facts are relatively dense they will be presented in more depth in relation to each issue rather than being repeated here.)

It will be argued that he presented a claim that was destined to fail. There are barriers to accepting the loss of a chance argument in itself, but even more so in a case such as this where the ‘chance’ has not been definitively lost. This is not because the claimant’s ‘chance’ was reduced to 20-30 percent,²⁷ but because the claimant was still alive, and until he either died or

²⁴ *Gregg v Scott* [2002] EWCA Civ 1471, [2003] Lloyd's Rep Med 105; [2005] UKHL 2, [2005] 2 AC 176.

²⁵ The reliability of these statistics will be considered below in addition to asking whether the loss of such a chance can constitute damage.

²⁶ The start date for measuring the ten year period is unclear. The medical evidence referred to ten years disease free survival which suggests it is measured from the date that the tumour has been eliminated. Lord Phillips thought that it probably started at the date of commencement of treatment (*Gregg v Scott* [2005] UKHL 2, [2005] 2 AC 176, [132]). At trial and in the Court of Appeal reference was made to the chance of five years’ disease-free survival but this seems to be because of the number of years that had elapsed between the negligence and the trial, so the medical definition of cure was still premised on ten years’ disease-free survival.

²⁷ The difficulty of giving an accurate figure will be addressed later.

survived disease-free for ten years it could not be said that this ‘chance’ had run its course. The problem seems to arise because the claimant used the loss of a chance argument to try to avoid the difficulty of proving a causal link between the negligence and the harm. But given that the claimant had not yet suffered the ‘outcome’ to which the chance related i.e. dying before ten years had elapsed, even if the evidence had favoured finding a causal link between the doctor’s negligence and the claimant’s probable death, he would still have been unable to claim for this because he had not yet suffered the loss. Effectively, faced with a case where it was going to be difficult to prove that the defendant’s negligence had caused loss to the claimant, he decided to advance a case based on loss of a chance, yet he made his loss of a chance argument more difficult/less palatable than it might otherwise have been by claiming for the loss of the chance of avoiding a harm that he had not yet, and may never, suffer. However, it seems that the claimant could have advanced a more modest claim based on traditional principles and that such a claim ought to have succeeded. Similarly to the discussion of *Hotson* in the previous section, the following analysis seeks to clear away any misconceptions surrounding how conventional negligence principles would apply to this case in order to be left with a clearer understanding of what problems remain that might be solved by taking a loss of chance approach to recovery.

2.1 Pain and suffering

During the nine months between the defendant’s negligence and the eventual diagnosis of the claimant’s illness the cancer had ‘gradually enlarged’.²⁸ Following the referral for a biopsy he was admitted to hospital with ‘acute and intense chest pain which was a result of the lymphoma having spread, in particular to the left pectoral region’.²⁹ The spread of the cancer is a personal

²⁸ *Gregg v Scott* [2002] EWCA Civ 1471, [2003] Lloyd's Rep Med 105, [1] (Latham LJ).

²⁹ *ibid* [2].

injury, and the claimant ought to have been able to recover for this physical injury and the consequential pain and suffering.³⁰

The next issue is the extent of the treatment that the claimant had to undergo. The claimant underwent chemotherapy, supplemented by radiotherapy. The tumour responded but not completely, so the claimant underwent a more aggressive treatment of high dose chemotherapy, with harvesting of stem cells. He suffered side effects from the treatment, and these were severe in the case of the high dose chemotherapy. He had to give up work, felt ill and had felt weak ever since. At first instance, Inglis J found that ‘it is possible to say on the basis of Professor Goldstone’s model that he would more probably than not have achieved complete remission with initial CHOP therapy and without high dose chemotherapy with stem cell harvesting’.³¹ Baroness Hale, despite rejecting the claim for loss of a chance, still acknowledged that this amounted to a finding that ‘the claimant would have achieved initial remission had he been treated earlier’.³² She ultimately considered that it was probable that the claimant would still have suffered the same setbacks and that the relapses would have happened anyway, but she did say:

Even on conventional principles, this does not necessarily mean that the claimant is not entitled to anything at all. The defendant is liable for any extra pain, suffering, loss of amenity, financial loss and loss of expectation of life which may have resulted from the delay. If, without the delay, the claimant would have achieved a longer gap before more radical treatment became necessary, then he should be entitled to damages to reflect the acceleration in his suffering. If the pain and suffering he would have suffered anyway was made worse by the anguish of knowing that his disease could have been detected earlier, then he should be compensated for that.³³

³⁰ *Gregg* (HL) (n26) Lord Hope, who ultimately would have allowed the claim for loss of a chance, recognised that he was in the minority and that the claim would therefore fail, but expressed his hope ‘that it was not too late for the pain and suffering which the appellant suffered due to the tumour’s enlargement...to be brought into account by way of an award of general damages’ [123].

³¹ [38] (Inglis J), cited in *Gregg* (HL) (n26) [202] (Baroness Hale).

³² *Gregg* (HL) (n26) [203].

³³ *Gregg* (HL) (n26) [206].

It is therefore clear that even if she thought that the delay had no effect on the overall prognosis for cure, so the spread of the cancer did not affect the final outcome of death/survival, she did think that the spread of the cancer was physical harm and that this had affected the treatment necessary which caused him greater pain and suffering. Again, he could have recovered for this if he had formulated his claim in terms of the spread of the cancer and the consequences of the spread.

2.2 Loss of life expectancy

The claimant argued that the defendant's negligence caused him to lose a chance of a cure by reducing this chance from 42% to 25%. Given that the medical definition of cure is ten years' disease-free survival, the claim for lost years was that the negligence had caused him the loss of a chance of ten years' disease-free survival. By incorporating the medical definition of cure into the claim for lost years the claimant seems to have made it easier for the court to reject his claim. This is because introducing a cut-off date of ten years adds a binary quality to the inquiry – he either will or will not live for ten years, and the delay in diagnosis either has or has not prevented him for living for ten years. If he had advanced a more modest yet conventional claim for loss of life expectancy, which is open-ended in nature, then it seems likely that he would have been able to establish that the delay in diagnosis did reduce his life expectancy. Admittedly his life expectancy was already quite short at the time of the negligence, but the spread of the cancer due to the delay reduced it even further.

Lord Phillips highlighted this problem which was created by the way the claim for lost years was formulated:

The conventional way of determining the effect of the injury on expectation of life is as follows. The court determines what the claimant's expectation of life would have been but for the injury. The court then determines what the claimant's expectation of life is

having regard to the effect of the injury. The difference between the two constitutes the “lost years”.³⁴

It is important to note that a conventional claim for loss of life expectancy is not the same as the ‘hook’ argument. The hook argument consists of arguing that once a claimant can point to some actionable damage that was caused by the defendant’s negligence, all other chances are recoverable as a matter of quantification. As discussed above in relation to *Hotson*, a claimant will actually only be compensated where the reduction in the chance of avoiding future illness or of making future gain is *as a consequence* of the actionable damage i.e. causally related to the actionable damage.³⁵ This is an area where the importance of distinguishing clearly between issues of damage, causation, and quantification, is very apparent yet often overlooked. Mr Gregg was unable to show that there was a causal link between the spread of the cancer and the final outcome of being cured (or failing to survive for ten years) because on the balance of probabilities he would not have survived ten years even with timely diagnosis. But if he had advanced a conventional claim for the physical harm constituted by the spread of the cancer, the additional pain and suffering caused by the spread, and the more aggressive treatment, then it would seem likely that having more extensive cancer would have reduced his life expectancy somewhat, even if his life expectancy was already short because he was suffering from cancer at the time of the negligence. If a claimant with an identical disease was shot and killed by a defendant, in valuing the lost years the court would assess his life expectancy with cancer at the time he was shot and compensate the lost years. So it is not making an impossible demand on the court to ask them to assess his life expectancy at the time of the negligence and his life expectancy nine months later when the cancer had spread.

Lord Phillips summarised the various ‘adverse events’ that Mr Gregg had suffered ‘beyond the initial development of his cancer’ as: i) the spread of the cancer to the pectoral region accompanied by acute pain; ii) high dose chemotherapy with harvesting of stem cells; iii) relapse

³⁴ *Gregg* (HL) (n26) [131].

³⁵ See text to (n13).

when a tumour developed in the right axilla accompanied by chemotherapy; iv) psychiatric distress on being told that the relapse meant that he would die; v) further suspected relapse with additional chemotherapy.³⁶ He correctly explained that ‘the chance that the delay in commencing his treatment has caused each of these adverse events is not the same’.³⁷ This highlights the need for claimants to be careful and specific in how they detail the loss for which they are claiming, and also the need for a claimant to be realistic about what he can prove. Given that corrective justice requires that a defendant only pay for that loss that he caused through his negligence, a claimant will only recover if he can prove the causal link between the negligence and the damage. It is surely preferable for a claimant to receive modest compensation for that portion of the loss the defendant can be proved to have caused, rather than to attempt to combine the losses into one claim for the loss of a chance of a better outcome. Lord Phillips concluded that

On balance of probability I suspect that one is now in a position to conclude that the delay in commencing Mr Gregg’s treatment has not affected his prospect of being a survivor but has caused him all the other adverse events which I have set out above.³⁸

If Mr Gregg had advanced a conventional claim for these limited heads of damage it is therefore likely that he would have been successful.

3. Statistics, chance, and the balance of probabilities standard of proof

This chapter has illustrated that it is essential to have a clear understanding of the concepts of damage, causation, and quantification and their interrelationship before it is possible to analyse the loss of chance argument for proportional recovery. In light of this it is clear that if courts recognised that the loss of a chance of a better medical outcome could constitute the gist of a negligence action, this would represent a significant departure from traditional principles.

³⁶ *Gregg* (HL) (n26) [167].

³⁷ *Gregg* (HL) (n26) [168].

³⁸ *Gregg* (HL) (n26) [169].

The loss of chance argument has been raised by claimants because of difficulties they faced proving the causal link between the defendant's negligence and their loss. In both *Hotson* and *Gregg*, the claimant was already unwell and the defendant's negligence consisted of a failure to diagnose and treat that existing illness. Proof of causation is naturally complicated in this scenario because there is routinely a candidate cause other than the defendant's negligence – the court must determine whether the illness alone would have caused the harm even with careful diagnosis and treatment. Since the civil standard of proof is the balance of probabilities, and recovery in negligence law is 'all-or-nothing', it is often said that where the patient's pre-tort chance of a cure exceeded fifty percent the courts regard him as having been destined to be cured, but if his chance of cure was below fifty percent the courts regard him as being doomed to suffer the harmful outcome even with careful treatment so the doctor's negligence is deemed not to be a cause and the claimant receives nothing.³⁹ In this section it will be argued that this is based on a flawed understanding of the balance of probabilities standard of proof, and on an overly simplistic reading of the statistical information that informs the assessment of the 'chance'. The balance of probabilities relates to the degree of belief the judge must form in the fact of causation,⁴⁰ and statistical evidence is just one element in informing a rational belief. The statistical evidence should be personalised to the individual as far as possible, and in order to form rational belief based on such evidence the judge must also be persuaded of the reliability of the evidence. The judge's degree of belief is not easily quantified as a percentage because it involves a qualitative assessment of the available evidence, and it would be wrong to think that statistics are somehow more useful or objective because they express probabilities. Understanding that the balance of probabilities relates to the judge's degree of belief highlights

³⁹ See Allan Beever, 'Gregg v Scott and loss of a chance' (2005) UQLJ 201, 201. See also Stapleton 'Loss of the chance of cure' (n10) 997; Sarah Green, 'Coherence of medical negligence cases: a game of doctors and purses' (2006) 14 Med LR 1, 3.

⁴⁰ See Richard Wright, 'Proving Causation: Probability versus Belief' in Richard Goldberg (ed) *Perspectives on Causation* (Hart Publishing 2011) 205-212.

the error of proportionate recovery, which would actually equate to discounting damages to reflect the judge's doubt as to causation.

3.1 What must the claimant prove? What is the problem of proof?

Lord Nicholls' willingness in *Gregg* to accept the loss of chance argument seems to reflect a clear desire to help a claimant overcome evidentiary difficulties in the medical negligence context. He explained that 'in cases of medical negligence assessment of a patient's loss may be hampered, to greater or lesser extent, by one crucial fact being unknown and unknowable: how the particular patient would have responded to proper treatment at the right time'.⁴¹ That, however, is the nature of a counterfactual inquiry, and it is addressed by using the balance of probabilities standard of proof. It is true that the existing disease makes it harder for a court to determine what would have happened to the claimant. In other contexts a claimant could have expected to carry on unharmed unless something unusual occurs, such as a defendant's negligence; in cases such as these where the claimant was already unwell then he would have expected to suffer harm unless the doctor could cure his illness. The loss of chance argument is therefore a solution to help claimants overcome the evidentiary difficulty, as Lord Nicholls makes plain:

In suitable cases courts are prepared to adapt their process so as to leap an evidentiary gap when overall fairness plainly so requires. *Fairchild v Glenhaven Funeral Services Ltd* is a recent illustration of this in a different context. In the present context use of statistics for the purposes of evaluating a lost chance makes good sense.⁴²

But although proof of causation is more complicated in these cases than in many straightforward scenarios, it is not afflicted with the same kind of 'evidentiary gap' as existed in *Fairchild*. Proof of causation was impossible in *Fairchild*. By contrast, in *Hotson* and in *Gregg* it was merely difficult. These are not cases where the only available evidence was statistical; there was evidence that could be personalised to the individual claimants. This section will explain the conventional

⁴¹ *Gregg* (HL) (n26) [27].

⁴² *Gregg* (HL) (n26) [31].

approach to what the claimant should prove, and the standard to which he should prove it, and apply this to the evidence available in the cases to show that it was appropriate to insist on traditional principles.

3.1.1 Damage

As argued in the previous chapters, the role of the causation inquiry in negligence is limited and is defined by the doctrines of damage and breach of duty. It is especially important to define the damage that forms the gist of the action from the outset because the negligence inquiry is concerned to determine whether the defendant's negligence was a cause of this damage, rather than with determining more generally what would have happened absent the negligence.

This is particularly relevant to the analysis of *Gregg* because the outcome of the claimant's treatment was still prospective so, apart from the limited heads of damage noted earlier, he had not suffered the harmful outcome to which the 'chance' related. This means that even if the claimant's pre-tort 'chance' of recovery had exceeded fifty percent he would not have had a successful claim on traditional principles unless he could prove that he had suffered actionable damage i.e. the harmful physical outcome. This means that Lord Nicholls was wrong to suggest that '[t]he patient could recover damages if his initial prospects of recovery had been more than 50%'.⁴³ The causation question is not the open-ended question of 'what position would the claimant be in if the defendant had acted carefully?', but the focused questions of 'has the claimant suffered actionable damage and was the defendant's negligence a cause of that damage?'

3.1.2 Causation

It is preferable to use the NESS test rather than the but-for test to address the causal link because it enables the court to pinpoint more accurately the issues of proof that arise.

⁴³ *Gregg* (HL) (n26) [2].

The advantages of the NESS test for causation are clear in *Hotson* because the NESS test treats each cause as having equal significance since each was equally necessary for the sufficiency of the set of conditions which resulted in the loss. As explained previously, avascular necrosis occurs when insufficient blood vessels remain intact to keep the epiphysis alive. When the claimant first went to hospital some vessels were damaged but some blood vessels remained intact although distorted. The delay in commencing treatment meant that already ruptured blood vessels continued to bleed into the joint thus increasing pressure in the joint and blocking the intact blood vessels. The delay in diagnosing and treating the injury therefore caused damage to those blood vessels that had remained and sealed the claimant's fate by turning his injury into an inevitability. When applying the NESS test the 'sufficient set' for avascular necrosis is 'insufficient blood vessels intact to keep the epiphysis alive'. The relevant question is whether the blood vessels damaged due to the delay were necessary for the sufficiency of this set. In other words, when the claimant first went to hospital were there sufficient blood vessels remaining intact for the epiphysis to stay alive?

Lord Bridge was quite clear that this was a problem of individual proof and did not involve any deeper problem that would necessitate resort to statistics. He said '[i]n some cases, perhaps particularly medical negligence cases, causation may be so shrouded in mystery that the court can only measure statistical chances. But that was not so here. On the evidence there was a clear conflict as to what had caused the avascular necrosis'.⁴⁴ In other words, if the problem was more akin to the evidentiary gap cases where there was a lack of understanding of the causal process and the medical evidence could only furnish statistical assessments of risk then there may be a reason for considering the loss of chance approach. But in this case the causal process was understood, the problem was simply one of proof in the individual case and the claimant had not managed to persuade the judge on the balance of probabilities.

Lord Nicholls sought distinguish the facts of *Gregg* from those of *Hotson* because there was:

⁴⁴ *Hotson* (HL) (n4) 782.

...no significant uncertainty about what would have happened to Stephen Hotson's leg if treated promptly, once his condition at the time of the negligence has been determined on the usual probability basis... Identifying Mr Gregg's condition when he first visited Dr Scott did not provide an answer to the crucial question of what would have happened if there had been no negligence. There was considerable medical uncertainty about what the outcome would have been had Mr Gregg received appropriate treatment nine months earlier.⁴⁵

The line that he draws between the degree and nature of the uncertainty in *Hotson* and in *Gregg* is not a clear one. Yet regardless of whether proportional recovery would be an appropriate exceptional solution, the 'evidentiary gap' in *Gregg* is not as deep as Lord Nicholls suggests. He rightly observed that statistical evidence alone is unable to identify the cause in a particular case, and statistical probabilities of over/under fifty percent do not translate directly into proof on the balance of probabilities. But, as the next section will illustrate, statistical evidence can have an important role in convincing a court that a causal link is more probable than not in a particular case where the evidence itself is reliable and where it can be personalised sufficiently to the claimant's case. Medical understanding of the aetiology of cancer may be incomplete but a range of risk factors are known so the statistical information may be personalised taking into account the characteristics of the individual such as age, general health, spread of the cancer etc. Indeed, Lord Phillips took a more rounded approach to the statistical evidence, examining its reliability and making some, limited, attempt to personalise it. The focus of the next section is on the role that statistical evidence can play in informing the balance of probabilities standard of proof because the courts need to understand this fully before they can identify accurately the problems that are raised by claims such as *Hotson* and *Gregg*.

⁴⁵ *Gregg* (HL) (n26) [38].

3.2 The balance of probabilities standard of proof

It is important to remember that the balance of probabilities standard of proof does not refer to ‘probability’ in the statistical sense, but refers to the degree of belief that must be established in relation to the fact of causation. It is a qualitative assessment that may be informed by statistical or epidemiological evidence but which ultimately must also incorporate an assessment of the reliability of the evidence and of how far it can be personalised to the individual claimant.⁴⁶

The balance of probabilities is often treated as a statistical tool, where causation is proven as soon as the 51% threshold is passed. In *Gregg*, Lord Nicholls began his speech by stating that ‘[t]he patient could recover damages if his initial prospects of recovery had been more than 50%. But because they were less than 50% he can recover nothing’.⁴⁷ This is also common in academic commentary, for example Beever has suggested that ‘[t]he problem facing the plaintiff was that, before the defendant’s negligence, his chance of being cured was only 42%. This meant that, on the balance of probabilities, the defendant did not deprive the plaintiff of a cure’.⁴⁸ Yet as Lord Nicholls observed:

Statistical evidence, however, is not strictly a guide to what would have happened in one particular case. Statistics record retrospectively what happened to other patients in more or less comparable situations. They reveal trends of outcome. They are general in nature. The different way other patients responded in a similar position says nothing about how the claimant would have responded.⁴⁹

He continued to ask,

Who can know whether Mr Gregg was in the 58% non-survivor category or the 42% survivor category? There was no evidence, peculiar to him or his circumstances, enabling

⁴⁶ Epidemiology is the branch of medicine that studies the incidence and causes of disease in populations.

⁴⁷ *Gregg* (HL) (n26) [2].

⁴⁸ Beever, ‘*Gregg v Scott*’ (n39) 201. See also Stapleton ‘Loss of the chance of cure’ (n10) 997; Green, ‘Coherence of medical negligence cases’ (n39) 3.

⁴⁹ *Gregg* (HL) (n26) [28].

anyone to say whether on balance of probability he was in the former group or the latter group.⁵⁰

It will be argued later that there was evidence enabling the court to say which group the claimant belonged to, but it is important to note that Lord Nicholls is right to say that raw statistics do not tell us anything about the individual. This means that if the raw statistics were reversed and 58% of patients would have survived, the court would still be unable to say whether the claimant fell into the survivor or non-survivor category if there was no information enabling them to personalise the statistics.

This tendency to map the patient's statistical chance directly onto the balance of probabilities is perhaps more common in cases of medical negligence because a doctor explaining a patient's prognosis will routinely refer to the statistical likelihood of recovery. A purely statistical approach may mean that where the patient's chance of recovering is very close to 50 percent the case will be seen as being very close to the margins. But once he has suffered the harmful outcome then it is a question of fact whether the condition was treatable, and the balance of probabilities is a matter of belief in this fact and is designed to deal with the uncertainty – the court does not need to be convinced that the injury was certainly treatable, it needs only be convinced that it is more likely than not that it was treatable. It is essential to understand the role that statistical evidence can play in informing the balance of probabilities standard of proof.

The balance of probabilities is a standard of proof or persuasion, so it is a 'belief probability'. Barnes explains that belief probability refers to 'the credibility – the believability – of the evidence in support of a party's factual claims'.⁵¹ He distinguishes this from 'fact probability' and 'sampling error probability'.⁵² Fact probability refers to the likelihood of a causal relationship. One aspect of this is the 'risk ratio' or 'relative risk':

⁵⁰ *Gregg* (HL) (n26) [29].

⁵¹ David W Barnes, 'Too Many Probabilities: Statistical Evidence of Tort Causation' (2001) 64 Law and Contemp. Probs 191, 192.

⁵² See also Steve Gold, 'Causation in Toxic Torts: Burdens of Proof, Standards of Persuasion, and Statistical Evidence' (1986) 96 Yale LJ 376.

Risk ratios measure the percentage change in the incidence of a specified harm, such as a disease. A risk ratio compares a background rate, where the stimulus in question is not present, to the rate that obtains when the stimulus is present. For example, in a routine tort case alleging that a negligent failure to light a stairway caused a fall, a risk ratio might compare the incidence of falling down stairs when the stairs are well-lit to the incidence of falling when the stairs are unlit...A risk ratio greater than one indicates that risks are increased. For instance, risk ratios of 1.5 and 3 indicate that the stimulus (for example, lack of lighting) increases the risk of falling by 50% and 200% respectively.⁵³

The ‘sampling error probability’ refers to ‘a statistical property of data underlying evidence offered to prove a relevant fact’,⁵⁴ so it aids the assessment of the reliability of a particular fact probability. Barnes explains that:

Even when a sample is composed of randomly chosen subjects, those subjects may not represent accurately the population. That possibility means that any observed statistical relationship between acts like the defendant’s and harms like the plaintiff’s revealed by a study of a sample may be due to the happenstance of having drawn randomly an atypical sample.⁵⁵

Barnes explains that in all studies based on a sample of the population there will be a sampling error and that this ‘is not an error in the design of the sample. Indeed, it is not an error attributable to any person. It is an unavoidable property of inferential statistics, the process of estimating attributes of a population by examining a sample’.⁵⁶ The sampling error probability is expressed as a ‘p-value’ ranging from 0 to 1.00, and a p-value closer to 0 indicates a smaller probability that error is due to the make-up of the sample. The greatest sampling error probability that is accepted in science is five percent, meaning that on the basis of sampling error

⁵³ Barnes (n51) 193.

⁵⁴ ibid.

⁵⁵ ibid.

⁵⁶ ibid 198.

there must be at least a 95 percent chance that the relationship is causal and not due to the chance that the sample is unrepresentative of the population. Furthermore:

A sampling error probability may as easily be calculated from a poorly designed study as from a randomized, controlled, double-masked study. The credibility of that probability and any fact probability derived from that study, however, depends on the quality of the design study.⁵⁷

In other words, '[t]he p-value does not measure the probability that the design was perfect; rather, it assumes the design was perfect'.⁵⁸

This means that in order for the court to form a rational belief in causation based on statistical or epidemiological evidence the reliability of the study must also be evaluated by addressing factors such as the experimental design and measurement. There are a range of criteria available for assessing the reliability of an epidemiological study and the statistics to which it gives rise. Barnes explains that the 'well-known gold standard for experimental design is the randomized, controlled, double-masked study'.⁵⁹

After the quality of a study has been established, epidemiologists then need to test the reliability of the causal inferences that can be drawn from the data. Miller has explained, '[i]n the UK, the nine criteria articulated by Professor Bradford Hill are perhaps the best known'⁶⁰ criteria to test causal inferences. These are: the strength of association, how high is the relative risk?; consistency, have the results been replicated elsewhere?; specificity of cause and of effect, does the 'cause' produce only one effect? Does the 'effect' have only one cause?; temporality, the effect must follow the cause; biological gradient, is a 'dose relationship' identifiable?; plausibility, is it consistent with prevailing biological knowledge?; coherence, does it conflict with any known biology of the disease?; experiment, does the frequency of the effect fall when the 'cause' is

⁵⁷ ibid 200.

⁵⁸ ibid 204.

⁵⁹ ibid 200.

⁶⁰ Chris Miller, 'Causation in personal injury: legal or epidemiological common sense?' (2006) 26 LS 544, 545.

removed?; analogy, is the cause-effect relationship of any comparable disease accepted?⁶¹ These factors relate to the question of whether there is a general causal relationship between the ‘cause’ and ‘effect’ i.e. the harmful agent for which the defendant is responsible and the type of illness suffered by the claimant. Miller notes that ‘Hill took pains not to exaggerate the claims of his criteria: “None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a sine qua non”’.⁶² While none of the factors alone is determinative of the quality of the study, together they enable its reliability to be evaluated.

Furthermore, in order to form a rational belief about the causal relationship in an individual case, the court must assess the extent to which it is possible to extrapolate from the fact probability about general causation to causation in the individual case. This involves other factors, notably how closely matched the claimant is with the sample population and with the general population. So even if a study is reliable, the sampling error probability is not the same as a belief probability because it ‘would be divorced from the belief probability just where we need it – where we ask whether the particular defendant’s act was a necessary event in producing the particular plaintiff’s harm’.⁶³ All of these factors therefore affect our overall degree of belief in the fact that the particular defendant’s negligence was a cause of the particular claimant’s loss, the belief probability. This means that statistical evidence can provide a starting point for forming a rational belief in causation in an individual case, but it must be interrogated for reliability and for how far it is capable of informing us about the individual claimant.

It is important not only for lawyers to understand what the ‘balance of probabilities’ entails but also to ensure consistency in the discourse between lawyers and medical experts. The trial judge in *Hotson* noted that the claimant’s expert witness ‘speaks of likelihood as something involving a

⁶¹ *ibid* 548.

⁶² *ibid* 548. Citing A Bradford Hill, ‘The environment and disease: association or causation?’ (1965) 58 Proc R Soc Med 295.

⁶³ *ibid* 204.

less than 50% chance in contradistinction to a probability as denoting more than that.⁶⁴ In other words, the medical expert drew a significant distinction between something being ‘likely’ and it being ‘probable’. Given the importance of the ‘balance of probabilities’ in determining the outcome of a case it is essential that all parties understand the concept and use the term ‘probability’ consistently.

Indeed the pressure from the lawyers to express the chances in statistical terms seemed to go against the medical experts’ position in *Gregg* which was that a figure could not be put on the individual chance. With regard to the claimant’s pre-tort chance of cure, Inglis J noted:

Professor Goldstone was unable to put a percentage figure on that chance. It has to be taken as less than evens...It might be said that since neither he nor Dr Bunch put a figure on it, the court should not. I do not agree...The experts thought it possible that his individual prognosis had been reduced to less than 50% of what it would have been intrinsically at the outset.⁶⁵

So since the balance of probabilities involves a qualitative assessment of the evidence it is important not only to avoid the temptation to map statistical probabilities directly onto it but also to avoid pressuring expert witnesses to quantify the risk if they are unable to.

3.3 Proof in *Hotson*

The details of the decision at first instance in *Hotson* show clearly that the judge did undertake a detailed qualitative assessment of the conflicting evidence.⁶⁶ That evidence did not consist simply of conflicting statistical probabilities of avascular necrosis following a fall. The 25% ‘chance’ was not a quantitative assessment of the likelihood of harm, but was a numerical expression of the judge’s degree of belief in the fact that the injury could be avoided; it reflected his qualitative assessment of the evidence. He looked at how severe the displacement of the bone was, how

⁶⁴ *Hotson v East Berkshire HA* [1985] 3 All ER 167 (QBD), 173.

⁶⁵ Para [51] (cited in *Gregg* (HL) (n26), [164] (Lord Phillips)).

⁶⁶ This is set out in the All England Law Reports (n64) but unfortunately is omitted from the Weekly Law Reports.

easily the fracture was reduced, the different interpretations of the x-rays, disagreement over whether the blood vessels would have continued to bleed and build up pressure on those remaining intact or would have sealed themselves, a paper from a medical journal that considered different treatment options, and a paper that analysed nine cases of fracture separation of the epiphysis in children. He found that there were internal inconsistencies in the evidence of the claimant's expert witness as well as between the two conflicting experts. On one issue he said 'I am not *persuaded* by the published material before me...Rather I am *on balance persuaded* to the contrary view'.⁶⁷ Eventually concluding:

I regret that I found certain parts of the evidence of both experts, highly qualified and experienced although they both undoubtedly are, difficult to accept, either as a result of internal inconsistency within their evidence or because of what seemed to be an intrinsic want of logic in some particular expressed view...In the result I find myself unattracted to, and finally unable to accept, either of the competing extreme views.⁶⁸

This was the basis for his finding that it was 'possible but improbable' that the fall had 'left intact sufficient vessels to keep the epiphysis alive'.⁶⁹ Whilst he expressed the 'chance' as a percentage, it is clear from the language used throughout the judgment that this 25% 'chance' was not a statement of how likely it was, given the number of blood vessels remaining intact, that the claimant could be cured, but a statement of how likely it was that enough vessels remained intact for him to be cured. It therefore reflected his degree of belief in the proposition that enough blood vessels remained intact for the epiphysis to be kept alive. It was expressed numerically to allow the calculation of damages at the valuation stage. Normally, if approaching this as a conventional claim for the physical harm, the judge would not need to put a figure on the extent of his doubt. The balance of probabilities merely requires him to find, as he did, that on balance he was persuaded that insufficient vessels remained intact. But because he was willing to

⁶⁷ *Hotson* (QBD) (n64) 173 (emphasis added).

⁶⁸ *ibid.*

⁶⁹ *ibid* 171.

approach it as a loss of a chance case, he has quantified the extent of the remaining doubt in his mind as being 25% so that he could then calculate a proportionate award of damages - this is in no way the same as saying that the doctor created a 25% risk, or caused the claimant to loss a 25% chance, or contributed 25% of the chance of harm.

3.4 Proof in *Gregg*

In *Gregg* the expert evidence at first instance consisted of a joint report by a consultant haematologist Professor Goldstone (instructed by Mr Gregg) and Dr Bunch (instructed by Dr Scott) concerning the claimant's disease and the effect of the delay in diagnosis, and oral evidence by Professor Goldstone alone. In the period between the claimant's pleading of the particulars of the claim in April 2000 based on Professor Goldstone's evidence, and the writing of the joint report in 2001, Professor Goldstone became aware of a paper that had been published in 1999 by the American Society of Hematology, written by Falini et al (the 'Falini paper') which suggested that the particular type of lymphoma could be further divided into two sub-categories, ALK positive and ALK negative. Those who fell into the ALK negative subset had a significantly worse prognosis than those in the ALK positive category. The claimant's lymphoma was ALK negative. The joint report by Professor Goldstone and Dr Bunch took account of the Falini paper. Only Professor Goldstone gave oral evidence and developed a 'working example' of the fate of a cohort of 100 ALK negative patients based on the evidence in the joint report. Latham LJ noted that '[t]he assumption was that the cohort consisted of patients with the same stage of disease as that from which the appellant suffered at the time when treatment should have been commenced'.⁷⁰

The working example showed:

Of those with initial treatment by CHOP chemotherapy with or without field radiotherapy 55 will achieve complete remission. 45 will not achieve complete remission,

⁷⁰ *Gregg* (CA) (n28) [14].

and of those 41 will then die. Four who did not achieve complete remission immediately will be brought to achieve it by further treatment of various kinds. Thus of the initial 100 59 manage to achieve complete remission. Of those 35 do not relapse. They are described by Professor Goldstone as the core group of survivors. 24, however, do relapse (usually, if they are going to, within two years or so of achieving remission). Of those 24 half, a further 12, will not be responsive to further treatment and will die. The remaining 12 will be responsive to further treatment, typically high dose chemotherapy with stem cell harvesting such as the claimant himself went through. Of those 12 half, a further six, will not relapse again and will become survivors. Of the remaining six who do relapse again, only one will survive. The number of survivors from the 100 who started out will therefore be 42.⁷¹

During the period of delay following the negligence the cancer upstaged from stage 1 (presence in the lymph nodes) to stage 1E (spread to other tissues outside the lymph nodes). There appears not to be a similar working example for a cohort of 100 ALK negative patients at stage 1E,⁷² but the expert evidence in the joint report was that ‘it was “quite possible” that his individual prognosis had been reduced to less than 50% of what it would have been intrinsically at the outset’.⁷³ At the time when the disease was actually diagnosed and treated the likelihood of cure was ‘no more than 10-15%’,⁷⁴ but at the time of trial ‘[h]is actual prognosis they considered to be dependent now more on the fact that he had survived for three years’ which meant that his chances of ten years survival ‘had increased from a statistical 10% to 20-30%’.⁷⁵ Thus ‘the Statement of Facts and Issues agreed by the parties included a statement that the judge found that the effects of the delay in commencing Mr Gregg’s treatment meant that his “chances of

⁷¹ Findings at first instance, cited in *Gregg* (HL) (n26) [146] (Lord Phillips).

⁷² As will be discussed later, this is because the working model was actually based on patients at all stages of the disease (see text to (n82)).

⁷³ *Gregg* (CA) (n28) [13] (Latham LJ).

⁷⁴ *ibid*.

⁷⁵ *ibid* [15].

long-term survival fell as a result of the negligence from 42% (on initial presentation to [Dr Scott]) to 25% (as at the date of trial)⁷⁶.

As noted at the start of this section, Mr Gregg had not suffered the harmful outcome to which this ‘chance’ related since he was still alive. Had he died, the causation question would be whether the negligent delay in diagnosis was a cause of his death, so it should be established on the balance of probabilities that his disease could have been cured. There are two areas of concern in this case: first, the reliability of the evidence; second, the court’s failure to personalise the statistical evidence to the claimant.

3.4.1 Reliability of the evidence

Lord Phillips’ examination of Professor Goldstone’s evidence led him to conclude that it was ‘a very inadequate tool for assessing the effect of the delay in treatment on Mr Gregg’s progress and prognosis’.⁷⁷ McIvor has identified a number of concerns regarding Professor Goldstone’s evidence which affect the degree of belief that can be placed in the statistics he presented to the court.⁷⁸ She notes that as a consultant haematologist his expertise is medical rather than statistical or epidemiological so whilst skilled in diagnosis and treatment of certain illnesses, he is less skilled in the interpretation of data. She further highlights that he had initially suggested that the statistical probability of survival was 84% because he had not taken into account the paper by Falini et al which showed that ALK negative patients had a much lower chance of survival than ALK positive patients. This paper had been available at the time of writing of his first report so its omission casts doubt on the quality of his evidence.⁷⁹ The Falini paper itself was based on a very small sample consisting of only 88 patients, 53 of whom were ALK positive and only 25 of

⁷⁶ *Gregg (HL)* (n26) [147] (Lord Phillips).

⁷⁷ *Gregg (HL)* (n26) [157].

⁷⁸ Claire McIvor, ‘The Use of Epidemiological Evidence in UK Tort Law’ in S Loue (ed.) *Global Perspectives in Forensic Epidemiology* (forthcoming).

⁷⁹ *ibid.*

whom were ALK negative.⁸⁰ Given the small sample size the accuracy of this data is probably limited. Furthermore, the ‘working example’ that Professor Goldstone devised to predict the likely fate of 100 ALK negative patients seems to be based on his own experience. Presumably given that he had only recently become aware of the Falini paper, and therefore was only recently averted to the reduced survival rate of ALK negative patients, it can only have been a rough model. This does not suggest that the statistics he produced are especially reliable, it suggests that they are the best that he could do based on his clinical experience and that a different expert could have given different statistics.

3.4.2 Personalising the evidence

Even if this evidence was reliable, in order to form a rational belief as to whether the claimant’s disease could have been successfully treated it is essential to personalise it to the individual as far as possible. Professor Goldstone had made some attempt to personalise the statistics to Mr Gregg’s post-tort chances in light of the progress of his treatment, suggesting that the likelihood of survival had increased from 10-15% at the time of diagnosis to 20-30% at the time of trial. But there was no corresponding attempt to personalise the statistics in relation to his pre-tort chance of survival despite there being information on which this attempt could have been based. Stapleton explains this argument:

The unpersonalized estimate of [Mr Gregg’s pre-tort chance of being cured] was 42%. But by the time of trial we had an important piece of information about the actual experience of this particular claimant after the commencement of treatment, which could have been used to “personalize” that estimate. By the trial we knew that, even after breach, Mr Gregg was able to achieve complete remission in 1996. Does this not show that, *a fortiori*, had there been no breach, Mr Gregg would have been one of the 59 out of

⁸⁰ *Gregg* (CA) (n28) [11].

100 who initially achieved complete remission? A member of *that* group had a pre-tort chance of cure of $42/59 = 71\%$, not 42%.⁸¹

It was noted above that it was assumed that Professor Goldstone's working example applied to a cohort of 100 patients whose disease was at the same stage as the claimant's pre-tort disease, and that he did not provide a similar working example for 100 patients whose disease was at the more advanced stage that Mr Gregg's cancer had reached when treatment commenced. It is thus unclear where the assessment of the claimant's post-tort likelihood of cure at 10%-15% was derived from. Furthermore, Lord Phillips questioned this assumption, explaining '[w]hen describing Professor Goldstone's model, the judge stated: "The 100 patients in the worked example include all ages, and also people with other unrevealed personal characteristics, *one of which is the stage of the disease at diagnosis*".⁸² This means that the court could have made some attempt to personalise the statistics by considering the claimant's age, and the stage of his disease. Surely, given the early stage of his disease he was more likely to belong to the survivor category than somebody whose disease is more advanced?

So in any negligence action the court must seek to personalise the evidence to the individual claimant in order to form a rational belief in the likely causes of the damage, and should also evaluate the reliability of the evidence. Mr Gregg's claim would have been hindered because he had not suffered actionable damage in the form of the final outcome, and because the evidence pertaining to causation had not been personalised and in any case was not reliable. It seems unlikely that he would have been able to prove causation on the balance of probabilities even if he had suffered actionable damage, but there is nothing to suggest that this is a particularly unusual problem requiring an exceptional solution. The claimant's argument was that the application of conventional principles resulted in an unfair failure of his claim, so exceptionally fairness demanded that the loss of a chance of a better outcome should be substituted as the gist of the action. This section has demonstrated that there were no inherent obstacles to proof of

⁸¹ Stapleton, 'Loss of the chance of cure' (n10) 1000.

⁸² *Gregg* (HL) (n26) [148].

causation that would have made rejection of his claim ‘unfair’, it is simply that the evidence in his case was weak, so there seems to be no reason why ‘fairness’ should have a claim on the law in preference to corrective justice. The next section turns to the notion of ‘chance’ itself to consider whether it could form the gist of a negligence action if fairness so required.

4. ‘Loss of a chance’ as damage

In this section it will be argued that in the context of medical negligence, specifically the misdiagnosis or mistreatment of an existing illness, the loss of a chance ought to be recognised as damage that can form the gist of a negligence action. Instead of approaching the question of loss of a chance as a way of assisting a claimant who faces difficulties of proof of causation, however, the issue will be approached from the perspective of whether a lost chance can constitute actionable damage. This means that ‘chance’ will be examined in its own right, rather than as leading to recovery of damages that are proportional to the physical harm to which the chance relates. In other words, it will be argued that a lost chance is a distinct form of damage that ought to have a place in the tariff set out in the Judicial College’s *Guidelines for the Assessment of General Damages in Personal Injury Cases*.⁸³ This will therefore continue to develop the argument that the issues of damage, causation and quantification are three distinct issues, and that clarity in the law can only be achieved by having a full understanding of the content and role of each concept and the interrelationship of the three.

This section will begin by addressing the concept of chance in order to demonstrate that, in the limited context of medical misdiagnosis/mistreatment of existing illness, a ‘chance’ is something of value that can be lost and ought to be recognised as being an actionable form of damage. There is a unique relationship between doctor and patient when treating an existing illness since the outcome of treatment is uncertain for both, meaning that both parties value the patient’s chance of being cured. As argued in Chapter One, these issues of interpersonal responsibility

⁸³ Judicial College, *Guidelines for the Assessment of General Damages in Personal Injury Cases* (11th ed., Oxford University Press 2012)

present in the doctor's duty of care should be allowed to shape the negligence inquiry so that loss of a chance can be recognised as damage. This is consistent with a corrective justice based approach to negligence which requires the defendant to repair wrongful loss that he caused. An important aspect of recognising loss of chance as actionable damage is the argument that the concept of the loss of a chance is distinct from the concept of an increase in risk, so recognition of claims for loss of a chance in this limited context would not have the expansionary effect that is often feared.⁸⁴ Finally the issue of valuing this chance will be addressed in order to show that since the chance is a distinct loss its value is independent of the final harm that the patient wishes to avoid. It will be clear that loss of a chance is not being presented as a solution to a problem of causation, and that this in turn will lend clarity to future causal problems in medical negligence.

4.1 Rejecting 'chance' conceived as a proportion of the physical outcome

The House of Lords in *Hotson* avoided the question of whether 'loss of a chance' was actionable because it was held that the patient did not have a chance of recovery when he first went to hospital. Instead, it was held, he either had sufficient blood vessels remaining undamaged and treatment would have succeeded, or insufficient blood vessels remained intact and he was doomed to develop avascular necrosis. Likewise, in *Gregg* it was held that the forty-two percent chance of recovery meant that in a cohort of 100 patients, 42 would recover and 58 would not.⁸⁵ The claimant either was one of the 42 or one of the 58, but he did not personally have a 'chance'. Such arguments are broadly based on the idea that the physical process of disease is 'deterministic' in nature. It is this issue of determinism of causal processes that will be examined in further detail. These cases were correctly decided. The fact that the claimants were unable to prove causation but put forward the loss of chance argument seeking to recover a proportion of the value of the relevant physical harm shows that they were effectively asking to be

⁸⁴ See for example Michael Jones, 'Proving Causation – Beyond the 'But For' Test' (2006) 22 PN 251

⁸⁵ Assuming these figures were based on reliable evidence.

compensated for the mere possibility that the defendant had caused them physical harm. As the previous section showed, the ‘chance’ was actually a quantification of the judge’s doubt over the causal link. This is inconsistent with corrective justice where a wrongdoer is required to repair a loss because he caused it through his wrongdoing.

Following the decision in *Hotson*, the concept of chance was elaborated in an article by Helen Reece,⁸⁶ in which she sought to reconcile that decision with the decision in *Chaplin v Hicks*.⁸⁷ In *Chaplin*, the claimant was a woman who had entered a beauty contest organised by the defendant. Entrants were narrowed down to a final group of 50 women who were then invited to an appointment with the defendant who would select 12 winners from this group. The claimant was unable to keep her appointment and sued the defendant for breach of contract for failing to give her a reasonable opportunity to present herself for an appointment. The defendant argued that damages could not be quantified, but the Court of Appeal held that since the claimant had roughly a one in four chance of winning (i.e. 12 out of 50 chance) she should be awarded 25 percent of the value of the prize that she would have received had she won the competition.

Reece argues that the distinction lies in the causal process itself which was ‘deterministic’ in *Hotson* but ‘(quasi) indeterministic’ in *Chaplin*. By ‘deterministic’ she means the ‘hypothetical chain of events is *fully determined by the events which have occurred*’.⁸⁸

This means that in a deterministic world, given sufficient knowledge the cause of anything could be discovered and the future could be predicted with certainty. In contrast, if a process is indeterministic it cannot be predicted even with unlimited knowledge. Reece explains that in an indeterministic process, probability is an objective concept. The likelihood of a future event occurring has an objective probability that is a property of the natural world. In contrast, in a

⁸⁶ Reece (n23).

⁸⁷ *Chaplin v Hicks* [1911] 2 KB 786 (CA).

⁸⁸ Reece (n23) 192. She provides the following explanation from Laplace: ‘We ought then to regard the present state of the universe as the effect of its anterior state and as the cause of the one which is to follow. Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective situation of the beings who compose it – an intelligence sufficiently vast to submit these data to analysis – it would embrace in the same formula the movements of the greatest bodies of the universe and those of the lightest atom; for it, nothing would be uncertain and the future, as the past, would be present to its eyes’ (Laplace, *Philosophical Essay on Probabilities* (1819, Springer-Verlag English translation 1995) 4).

deterministic process there is no objective probability of something occurring – either it will or it will not and theoretically this can be predicted with certainty. So when probability is used to describe the chance of such an event, the probability is not objective but epistemological; it is an expression of the likelihood of an event given the limited knowledge that is available. Reece explains, for example, that in an accident where a motorist drove into a pedestrian who then suffered broken bones, the bones broke because the force of the car was applied to them ‘and bones do not break in the same way on a random basis’.⁸⁹ Similarly, given sufficient knowledge it would be possible to predict with certainty whether or not a pedestrian’s bones would be broken in any particular case. So when road safety advertisements say that if a driver hits a child at 40mph there is an 80 percent chance that the child will die, but at 30mph there is an 80 percent chance that the child will live, this is an epistemological assessment of the ‘chance’ that the child will die, but in any particular case the outcome (life or death) would be determined by the precise conditions. Given sufficient knowledge, in the instant that a child is hit by the car it would be possible to say with certainty that this child will live or die. In contrast, the decay of a Uranium atom is random, and similarly the outcome of a lottery is random and cannot be predicted so the holder of a lottery ticket has an objective chance of winning.⁹⁰ In these indeterministic cases, Reece concludes, the balance of probabilities standard of proof simply cannot cope so the courts turn to a loss of chance approach which reflects the underlying existence of an objective probability.⁹¹ So she rejects the notion of ‘loss of a chance’ as damage unless there was an objective chance.

Reece argues that the causal process involved in *Hotson* was deterministic:

[T]here was a time in the past when the cause of the necrosis could have been determined. If the blood vessels had been examined after the fall, then it would have been humanly possible to decide whether or not the plaintiff would develop necrosis

⁸⁹ Reece (n23) 195.

⁹⁰ ibid.

⁹¹ Reece (n23) 204-5.

even if he were treated. At the time of the trial the cause was uncertain, but the uncertainty was epistemological not objective.⁹²

Since the claimant did not have an objective chance of recovery, Reece concludes that the ‘loss of a chance’ cannot form the gist of the negligence action. It is clear that once a claimant has suffered physical harm the causes of this are fixed because the physical process was deterministic, so there is not a ‘chance’ that the defendant was a cause there is merely uncertainty as to whether he was a cause. This means that the loss of chance argument, as it was presented in *Hotson* and in *Gregg*, cannot be accepted.

Understanding that the ‘chance’ in these cases was an epistemological probability explains why the patient’s ‘chance’ fluctuated in *Gregg*. The House of Lords was reluctant to accept the loss of chance argument in *Gregg* because of the apparent instability of the ‘chance’ since the statistical probability of the claimant being cured fluctuated over time as his treatment progressed. It is clear that this is more than simply a practical concern about the date at which the ‘chance’ should be measured if the claim had been allowed, but reflects the underlying fact that he did not have an objective chance. The ‘chance’ is an epistemological probability, meaning that it reflects the uncertainty over the outcome, so as the patient gets closer to achieving an outcome the uncertainty decreases so his ‘chance’ will get closer to 0% or 100%.

4.2 Epistemological chance as an independent form of damage

As argued above, once the claimant has suffered physical harm the concept of loss of a chance is generally rejected on the basis that a patient does not have a ‘chance’ as such. Either he was destined to recover, or he was destined not to, but he did not have a ‘chance’ of recovering. Rather, the ‘chance’ is an epistemological probability, that is it is an expression of our belief in the likelihood that the patient is one of the people destined to recover based on the limited knowledge available. The argument that will be developed here is that in the face of uncertainty

⁹² *ibid* 196.

at the time of diagnosis and treatment both the patient and the doctor treat the epistemological chance as though it were an objective chance. There is a unique relationship between doctor and patient when treating an existing illness since the outcome of treatment is uncertain for both, meaning that both parties value the patient's chance of being cured. These issues of interpersonal responsibility present in the doctor's duty of care mean that the law should adapt to the reality of that uncertainty rather than striving towards a philosophically correct approach.

It is important to remember at the outset of this section that what is under consideration is simply the question of whether the patient's chance of recovery is something that is capable of constituting loss. This should be kept separate in the mind from the question of whether the patient also suffers physical harm because this is not an attempt to assist a claimant who is facing difficulties of proof of causation. It should also be kept separate from the question of valuation of the chance when assessing damages.

4.2.1 Post-diagnosis factors that determine the outcome

Even if the physical process is deterministic the outcome of treatment will often be beyond the control of the doctor who cannot guarantee a cure, and this is particularly true when the outcome is affected by factors beyond his control that arise after the diagnosis. In other words, although the causal process may be deterministic, the outcome is not always determined at the time of diagnosis. Things such as the patient's diet and lifestyle may affect the course that the treatment takes, so the doctor regards the treatment as having a chance of success.

The extended time period for treatment of the disease in cases such as *Gregg* means that medical advances may be made over the course of the treatment that improve the prospects of cure as Lord Nicholls noted when he said '[n]ew and improved drugs and procedures make possible ever more alleviation of illnesses and injuries'.⁹³ In the Court of Appeal, Mance LJ explained:

⁹³ *Gregg* (HL) (n26) [23].

[T]he most obvious influencing factors are, one would suppose, internal to the claimant at the time of the negligence, however unknown or unknowable they may be; and they consist of the precise characteristics, development and spread of the cancerous cells at the time of the negligence as well as the claimant's precise physical characteristics and resistance. Other influencing factors may very well include subsequent events such as the particular medical treatment received, the patient's subsequent lifestyle and his or her, or indeed others', reaction to the stress inevitably incurred.⁹⁴

This illustrates that even if the causal process of cancer is deterministic, the outcome was not necessarily determined at the time of the negligence. If subsequent factors such as the patient's lifestyle and his response to stress are factors that determine the outcome then it is not possible to say whether, at the time of the negligence, Mr Gregg was fated to die or destined to recover. The possible impact of stress upon the course of his disease is especially relevant because the delay in diagnosis and consequent knowledge that his cancer had spread, which meant that he was statistically more likely to die, will have affected the patient's physical and mental state. This means that the negligence will have contributed to a factor (stress) which is one of the factors that determines whether or not he will be cured.

This suggests that a patient's physical state at the time of negligence was not, on its own, determinative of the outcome so a doctor could not even theoretically predict with certainty whether the treatment would lead to a cure. This inherent uncertainty means that he considers the patient to have a chance, and the law should reflect this reality in recognising a loss of that chance as damage. However this chapter seeks to make a more far-reaching argument that even where the outcome is determined at the time of the doctor's negligence, as in *Hotson*, at the time of the negligence both the patient and the doctor regard the prognosis as uncertain and therefore consider the patient to have a chance of recovery. This means that any patient whose illness is misdiagnosed or mistreated ought to be considered to have lost a chance if that delay affects the

⁹⁴ *Gregg* (CA) (n28) (Mance LJ). See also *Gregg* (HL) (n26) [80] (Lord Hoffmann).

medical assessment of the likelihood of cure. In this model the final physical harm is still seen as having a determinate cause which ought to be established on the balance of probabilities. The loss of chance argument for proportionate recovery has no place in this backwards-looking issue. However, if we situate ourselves at the time of the negligence, looking forward, the outcome is uncertain in the eyes of both the doctor and the patient, so both parties regard the patient as having a chance of cure at that time, and any damage to this chance is a distinct form of harm deserving of compensation. Since this harm is distinct from the final outcome its value is also independent of the final outcome rather than being valued as a proportion of the physical harm. Because it is an independent form of harm this also means that it would be actionable even if ultimately the harm materialises.

4.2.2 The grey-areas on the deterministic-indeterministic spectrum

In her article on loss of a chance, Reece adopted a working understanding of determinism and indeterminism which showed two extremes which enabled her to illustrate the different types of causal process, but she reached the more nuanced conclusion that ‘a continuum between determinism and indeterminism is more plausible than a clear division’.⁹⁵ It will be argued in this section that medical negligence cases involving misdiagnosis or mistreatment of existing illness do indeed fall somewhere in the middle of this spectrum. The physical causal process may well be deterministic, so once physical harm has occurred it is theoretically possible to discover the causes of this outcome. At the time of diagnosis it is also theoretically possible, with unlimited knowledge and understanding, to predict the outcome with certainty. But crucially at the time of diagnosis and treatment the practical reality is that both the claimant and the defendant doctor have only limited knowledge and understanding so they can only state the likelihood of a cure. Although they may be aware that the philosophical/scientific truth is that this is an epistemological probability, they both treat the patient as having a personal chance of cure. The

⁹⁵ Reece (n23) 206.

patient may adapt his lifestyle to minimise risk factors and improve his chance of cure, the focus of the doctor's duty of care is on treating the risk factors (i.e. the illness) in order to maximise the patient's chance of being cured. This characteristic of the doctor's duty of care means that the demands of interpersonal responsibility favour the recognition of this form of damage in order to achieve a coherent approach. The causal process is deterministic but in the context of this special relationship both parties treat it as being indeterministic, indeed the purpose of the defendant's duty of care is premised on it being indeterministic. This means that the damage to this epistemological 'chance' is a loss that ought to be repaired under corrective justice.

Reece acknowledges that it is difficult to conclude that an event is indeterministic because 'our inability to predict its occurrence could be either because the event is inherently unpredictable or because we have not yet found the complete set of necessary and sufficient causes'.⁹⁶ She is willing to characterise an event as 'quasi-indeterministic' i.e. to the best of our knowledge indeterministic:

if and only if it could not have been predicted at any point in the past, it cannot be predicted in the present even given unlimited time, resources and evidence, and we cannot imagine how it would become predictable in the future, even given the success of present research programmes.⁹⁷

This is clearly a very demanding definition. It sets a high threshold and means that most physical processes will be regarded as deterministic.

It is clear that a grey area exists if one compares the views of different authors regarding an event as familiar as the toss of a coin. Reece treats this as indeterministic, stating that '[t]ossing a coin does not cause the coin to land heads, but it does cause the chance that the coin will land heads'.⁹⁸ In contrast, Beever regards it as deterministic, explaining '[w]hen the coin is in the air there is a fact of the matter as to which side it will land on. There is, then, no objective chance

⁹⁶ ibid 194.

⁹⁷ ibid 194.

⁹⁸ ibid 206.

here whatsoever'.⁹⁹ This highlights the fact that where it is theoretically possible to predict an outcome, but impossible to predict it given normal knowledge or experience, it is common to treat an event as being indeterministic.

Reece's definition of determinism may be an acceptable theoretical and philosophical definition but it does not reflect the experience of the doctor and the patient. Even with his medical knowledge and understanding, the doctor is unable to predict the outcome with certainty, instead describing the patient as having a chance of recovery and then treating the illness. Providing appropriate treatment is the way the doctor finds out whether the patient could or could not be cured. As Lord Nicholls argued, the doctor regards the patient as having a chance of being cured, explaining:

the law should be exceedingly slow to disregard medical reality in the context of a legal duty whose very aim is to protect medical reality. In these cases the doctor's duty to act in the best interests of his patient involves maximising the patient's recovery prospects, and doing so whether the patient's prospects are good or not so good.¹⁰⁰

In a case such as *Hotson* the process is deterministic, and medical science has advanced enough to understand that it is the state of the blood vessels that determines whether the damage will occur. But, in any individual case, when the doctor commences treatment both he and the patient regard the outcome as involving a chance because they simply do not know whether sufficient blood vessels remain intact. Even with careful treatment the doctor cannot guarantee that the treatment will succeed or fail because there are factors beyond his control and because he does not know whether there are enough blood vessels remaining intact.

In a case such as *Gregg* it is even clearer that the doctor treats the patient as though he has a chance of recovering. When Mr Gregg was eventually referred to a specialist who diagnosed the

⁹⁹ Beever, '*Gregg v Scott*' (n39) 209.

¹⁰⁰ *Gregg* (HL) (n26) [42].

lump as cancer he was told that he had a 10-15percent chance of survival if treated.¹⁰¹ Even armed with knowledge of the extent and location of the cancer, and with understanding of factors that affect survival rates, the doctor was unable to say whether Mr Gregg was one of the 10-15 out of 100 who would survive or one of the 85-90 out of 100 who would die even with treatment.

Since this argument is based on the medical reality and shared experience of the doctor and patient, it would apply in all cases of misdiagnosis/mistreatment of existing illness. Since it is not an exceptional solution to a problem of proof of causation there would be no need to seek to circumscribe its application by reference to vague or arbitrary criteria. In contrast, the loss of chance argument for proportionate recovery would enable claimants to bypass the causation requirement, so its supporters have sought to limit its application by reference, for example to the degree of uncertainty in the disease. Jansen argues that ‘chance’ exists where there is ‘a genuine, unavoidable uncertainty, relative, of course, to standards of scientific knowledge’ so in his example the question of whether someone will survive a cardiac arrest ‘can only be answered in probabilities and is therefore best perceived in terms of chance and risk’.¹⁰² Similarly Lord Nicholls in *Gregg* suggested that a loss of chance approach to proportionate recovery was appropriate there but not in *Hotson* because the patient’s prospects in *Gregg* but not in *Hotson* were ‘fraught with a significant degree of medical uncertainty’.¹⁰³ Lord Nicholls acknowledged that this might be criticised for being ‘an imprecise boundary’,¹⁰⁴ and such criticism is arguably well-founded. Most cases of medical negligence will involve a degree of uncertainty surrounding causation and that is precisely why the law only requires proof on the balance of probabilities. If courts allowed proportionate recovery in some cases where they felt that the degree of uncertainty was great enough, decisions would essentially turn on whether the court felt that the

¹⁰¹ *Gregg* (CA) (n28) [13] (Latham LJ). This was his ‘chance’ at the time of diagnosis – by time of trial his chance was assessed at 20-30%.

¹⁰² Nils Jansen, ‘The Idea of a Lost Chance’ (1999) 19 OJLS 271, 279.

¹⁰³ *Gregg* (HL) (n26) [49].

¹⁰⁴ *ibid.*

claimant was deserving of help in avoiding difficulties of proof. In contrast, recognising that every patient who is already ill has a chance of being cured, and that this chance is something of value to him independently of the outcome (i.e. not leading to proportionate recovery for the physical harm) because at the time of treatment both the patient and the doctor see the outcome as being uncertain, then loss of chance would be recognised in every case where medical practice is to describe the patient as having a chance. When the concept of chance is analysed on its own terms as arising out of the considerations of interpersonal responsibility unique to the doctor/patient relationship there is no need to try to draw lines to limit it to being an exceptional approach.

4.2.3 When does a patient have a ‘chance’? Chance vs. risk.

It is still necessary to address the concern that loss of a chance is no different from an increase in risk. It will be argued that this concern arises because recovery for loss of a chance is usually presented as a solution to difficulties of proof of causation and as leading to proportionate recovery. This bears some resemblance to the application of the *Fairchild* principle in *Barker* which led to proportionate liability.¹⁰⁵ By considering chance independently from the outcome, and as arising because of unique characteristics of the doctor-patient relationship, it will be clear that chance is conceptually distinct from risk. This is because it arises in the limited context of medical misdiagnosis or mistreatment of existing illness where the shared uncertainty of the doctor and the patient means that ‘chance’ becomes a meaningful concept.

Jones has noted the apparent similarities between chance and risk:

Although their Lordships would no doubt protest that the facts of *Barker* are at some remove from *Gregg v Scott*, the acceptance that an increase in the risk of harm can constitute actionable damage (where the harm has actually materialised) indicates that there is no conceptual objection to a claim based on the loss of a chance of a better

¹⁰⁵ *Fairchild v Glenhaven Funeral Services (t/a GH Dovener & Son)* [2002] UKHL 22, [2003] 1 AC 32; *Barker v Corus (UK) plc* [2006] UKHL 20; [2006] 2 AC 572. These cases are addressed in Chapter Four.

medical outcome. The explanation for the different outcomes comes down to a choice as to which claimants, faced with causal uncertainty, deserve to be given a helping hand in the litigation lottery.¹⁰⁶

If loss of a chance did lead to proportionate recovery then it would be conceptually close to risk. The claimant in *Barker* suffered mesothelioma caused by asbestos inhalation but the ‘evidentiary gap’ in relation to the aetiology of that disease means that where the asbestos has come from a number of former employers it is impossible to say on the balance of probability that any individual source was a cause of the harm. Exceptionally the *Fairchild* test of ‘material contribution to the *risk* of harm’ applies. In *Barker* the House of Lords held that the *Fairchild* test should lead to several liability so that each defendant is liable only for the extent of his contribution to the overall risk. Given that the overall risk was a risk of the physical harm that the claimant suffered, this amounted to liability for a portion of the physical harm. This was a defendant-focused solution, the claimant was still notionally entitled to compensation for the full value of the physical harm (subject to contributory negligence) but each defendant was only liable for a portion of the total loss according to the extent of his contribution to the total risk of harm. The loss of chance approach proposed by the claimants in *Hotson* and in *Gregg* awards partial recovery for the claimant although the effect would be comparable for the defendant doctor who would compensate a proportion of the physical harm corresponding to his contribution to the overall risk.

Yet the risk-based approach was applied in *Barker* because of an evidentiary gap relating to the overall causal process, whereas the problem in *Hotson* and in *Gregg* was that despite the understanding of the causal process the claimant was unable to prove his individual case. In terms of the NESS test, the claimants in *Hotson* and in *Gregg* were unable to prove that the negligence was a *necessary* element of the sufficient set. With better evidence he would have been able to prove that the negligence was a cause of his loss. But in *Fairchild* and *Barker* the general

¹⁰⁶ Jones (n84) 268.

lack of understanding of the aetiology of the disease meant that one could never prove a causal link. The uncertainty there was what constitutes a *sufficient set* in the first place let alone requiring the claimant to prove that the negligence was *necessary* on this occasion.

The problems of proof are therefore very different in *Barker* and in the loss of chance cases, but if proportionate recovery was allowed in loss of chance cases it would effectively achieve the same outcome by holding the defendant liable for his contribution to the total risk. This would, therefore, have extended the boundaries of the *Fairchild* exception. Limits to the application of proportionate liability would have to be sought beyond corrective justice and in policy arguments. For example, Green considers whether the *Fairchild* exception could be applied in a case such as *Chester v Afshar* and beyond noting that *Fairchild* addressed the specific problem of the evidentiary gap, she turns to policy arguments to distinguish the cases, noting that

Fairchild was concerned with a claimant who had been injured by his employer, whilst that employer was acting solely in pursuit of commercial gain. This is not obviously comparable to a claimant who has been injured during a surgeon's attempt to ameliorate a naturally occurring injury.¹⁰⁷

In contrast, recognising that the chance of a cure is valued by the claimant and defendant independently of the physical outcome, is consistent with corrective justice because it simply recognises damage to this chance as a loss caused by the defendant. It does not lead to proportionate recovery for the physical outcome so it is conceptually distinct from the notion of contribution to risk.

4.2.3.1 When does a patient have a meaningful chance?

It is also important to note that a patient only has a meaningful chance when he is already unwell and the negligence relates to the treatment of this illness. It is the existing illness that creates the shared uncertainty for him and the doctor. In contrast, if the doctor introduces a new risk of

¹⁰⁷ Green, 'Coherence of medical negligence cases' (n39) 19.

harm then the situation is like any other case of negligence and it is not meaningful to talk of the claimant having a ‘chance’. This means that the approach to loss of chance suggested here would not alter the decision in *Wilsher v Essex Area Health Authority*.¹⁰⁸ In that case the claimant was a baby who had been born prematurely and later suffered blindness caused by a condition called retrolentalfibroplasia (‘RLF’). While in hospital he received negligent medical treatment when a catheter was inserted into a vein rather than an artery leading to incorrect measurement of the oxygen levels in his blood and the hospital therefore administering excess oxygen. Excessive oxygen in the blood is known to cause RLF, but due to his being born prematurely the baby also suffered four other conditions that each also carried a risk of RLF. He was unable to prove on the balance of probabilities that it was the excess oxygen rather than one of the naturally occurring conditions that caused his blindness and the House of Lords refused to apply the *McGhee* test of material contribution to the risk of harm in this case.¹⁰⁹ Since it was not possible to identify on the balance of probabilities which harmful agent (e.g. oxygen) had caused the illness, it was not possible to say that the defendant had contributed to the relevant risk factor.

Stapleton has argued:

[T]he pure loss-of-a-chance claim in *Gregg* might indirectly have put pressure on the decision on *Wilsher* which has been recently and unanimously approved by the Lords in *Fairchild*. At the moment after breach might baby Wilsher have had a loss of a chance claim of the sort contemplated by Lord Nicholls because his premature condition gave rise to a “significant medical uncertainty”? Even the adoption of the control mechanism, or “artificial hook”, of requiring that the breach had caused a physical change might not prevent such a pre-blindness claim being made by baby Wilsher, who could argue that the excessively oxygenated blood resulting from breach was such an “injury”.¹¹⁰

¹⁰⁸ [1988] AC 1074 (HL).

¹⁰⁹ The *McGhee* test was applied in *Fairchild* and is more commonly known as the *Fairchild* test today, but *Wilsher* was decided in the interval between those two decisions.

¹¹⁰ Stapleton, ‘Loss of the chance of cure’ (n10) 1003.

There are two issues to address here. The first relates to the ‘hook’ argument. This has already been discussed in detail to show that there would have to be a causal relationship between the physical ‘hook’ and the related chances. It is important to note that a physical change does not constitute damage until it makes the claimant worse off, so excessively oxygenated blood would not constitute a ‘hook’.¹¹¹

The second issue is that in *Wilsher* the doctor exposed the baby to a discrete risk, independent of the naturally occurring risks, so the baby should not be described as having a ‘chance’ of avoiding RLF that the doctor caused him to lose. The fact that the claimant could not show that it was oxygen that caused the harm meant that the House of Lords held that the defendant could not be shown to have contributed to the relevant risk, so the negligence had not made a ‘material contribution to the risk of harm’. Understanding this also explains why a patient may lose a chance of cure when a doctor fails to diagnose or treat his existing illness, but does not lose a chance when the doctor introduces a new risk.

Lunney suggests that *Gregg* would affect the analysis of *Wilsher* suggesting that ‘both *Wilsher* and *McGhee* are examples of attempts to deal with evidentiary uncertainty as to causation and could all be described as loss of chance cases. In *Wilsher*, it could be argued that the plaintiff lost the chance that he *would not* have contracted RLF from non-negligent exposure to excess oxygen’.¹¹²

This is surprising because in the introduction to his article, Lunney says:

The exact definition of ‘loss of chance’ is unclear, but it is generally accepted that it refers to the loss of an opportunity to obtain or receive a desired outcome; for example, recovery from a personal injury. Accordingly, where a patient receives negligent treatment in circumstances where, even if the treatment was properly carried out, that patient might not have been cured it is said that the patient has lost the chance of being cured as a result of the treatment.¹¹³

¹¹¹ *Rothwell v Chemical and Insulating Co Ltd* [2007] UKHL 39, [2007] 3 WLR 876.

¹¹² Mark Lunney, ‘What price a chance?’ (1995) 15 LS 1, 8.

¹¹³ *ibid* 1.

Although Lunney did not set this out as a definitive definition, the fact that he said that chance arises *in circumstances where, even if the treatment was properly carried out, that patient might not have been cured* implies that he sees chance as arising only where there is an existing risk and it is that risk that the negligence allows to develop into harm. Stapleton has said that:

[I]n *Wilsher* at least one source of the risk was *known* to have been an innocent source entirely unrelated to the defendant. If the loss of a chance claim had succeeded in *Gregg*, also a case involving a known unrelated unequivocally innocent source, this would have suggested that a moment after breach a future baby Wilsher would have a claim.¹¹⁴

But although the source of risk in *Gregg* was innocent it was not ‘entirely unrelated’. The cancer arose independently of the doctor, but it was not unrelated to the doctor’s negligence because it was this specific risk that the doctor was supposed to identify. In the approach to loss of chance being proposed in this chapter it is clear that a patient has a ‘chance’ of recovery when he is already unwell and the medical negligence consists of a failure to diagnose or treat that particular illness meaning that the likelihood of the patient recovering from that illness decreases. The concept of chance is meaningful in that context because the patient is already at risk, the doctor’s duty is focused on that specific risk, and both the doctor and the patient are faced with uncertainty regarding the patient’s prognosis.

If the doctor introduces the patient to a new risk through negligent mistreatment, such as adding excessive oxygen to the patient’s blood, then he might cause physical harm but he has not damaged the patient’s ‘chance’ of recovering because the patient did not have a meaningful ‘chance’ of avoiding that harm. To say that the patient has a chance of avoiding new illnesses is too general a conception of chance for it to be meaningful. In contrast it is meaningful to describe a patient who is already unwell as having a chance of recovering from that illness because his prospects are already damaged and uncertain.

¹¹⁴ Stapleton, ‘Loss of the chance of cure’ (n10) 1003.

A chance exists because the patient's illness creates an uncertainty that is specific to him rather than the general level of uncertainty to which background risks expose everyone. For example, in hospital there is a general background risk of contracting MRSA and every patient is exposed to this risk so it is not meaningful to describe an individual as having a 'chance' of avoiding it. But if a patient is already suffering from MRSA and the doctor negligently failed to diagnose it then the likelihood of that individual dying from it was already higher than the normal person and *his* prognosis was uncertain so he had a meaningful 'chance' of avoiding the harmful outcome.

4.2.4 'Loss' of a chance

It is important to remember that there must be a 'loss' of a chance and that the doctor's negligence must be shown, on the balance of probabilities, to be a cause of this loss.

When loss of chance is presented as a solution to the difficulty of proving causation, one objection is that it is not possible to say that the chance has been 'lost' unless it has been reduced to zero. In other words, it is only once the claimant has suffered physical harm that his chance has been definitively lost. In *Gregg*, Lord Phillips therefore left open the question of whether proportionate recovery for loss of a chance would be appropriate where physical harm had actually occurred, but on the facts of *Gregg* he considered it impossible to say whether a chance had been lost because the physical harm had not yet occurred. This problem simply does not arise in the approach proposed in this chapter where the concept of chance is meaningful because at the time of the diagnosis and treatment of the illness both the doctor and the patient regard the prognosis as being uncertain and therefore value the statistical assessment of the patient's chance of cure. This means that it is the fact of a reduction in the statistical chance that constitutes a 'loss' of a chance.

Lunney has argued that in the medical negligence context it is inappropriate to talk of 'loss of a chance' because the chance is not definitively lost. The patient still undergoes treatment and the remaining chance runs its course. For him this differs from cases such as *Chaplin* where the

claimant was definitively deprived of the opportunity to enter the beauty contest. Lunney explains this through the analogy of a racehorse injured by a driver's negligence on its way to a race. He says that if the accident prevents the horse from racing, then the negligence has 'deprived the horse of the chance to win the race [so] this would be a true lost chance case'. If, however, the horse suffers shock but is still able to run and finishes seventh there would be no claim for the loss of the chance of a better outcome: 'the horse has run the race and has lost, and either the accident caused the horse to lose or it did not. There is no question of chance; it is for the court to decide whether the negligence was responsible for the outcome (the horse losing the race)'.¹¹⁵

However arguably it is the fact that in the medical negligence context the patient *has* to run the chance with decreased odds that makes the chance so significant and valuable. In the context of a race, the horse stands to make a gain. In the medical negligence context the patient stands to suffer illness and possibly death. The owner of the horse has the choice whether or not to run the race with lower odds. The patient has no choice but to undergo treatment with a lower chance of recovery. Furthermore, the risk of the horse losing the race is a risk that is independent of the driver. In contrast, where a patient is already ill they have a chance of recovering and it is the purpose of the defendant doctor's duty of care to diagnose and treat that illness and protect the patient's chance of recovery.

Traditional negligence principles would still apply to this damage, so the claimant must prove on the balance of probabilities that the negligence caused the reduction in his chance of cure. In *Hotson* the judge found that the delay had allowed the bleeding to continue to build pressure in the joint and block those blood vessels that had remained intact, turning his injury into an inevitability. This meant that his physical state had worsened so his chance of cure was reduced by the delay caused by the negligence. In *Gregg* the claimant's cancer spread during the nine months delay so it was clear that his physical state had worsened and this was responsible for the

¹¹⁵ Lunney (n112) 5.

reduction in the statistical likelihood of cure. If the cancer had remained unchanged during that time then it would be hard to see that the delay caused any reduction in the chance of cure.

4.2.5 Valuing the chance

The final question is how to value this chance. In this approach the lost chance exists independently of the physical outcome and this has important consequences for the valuation of the chance because it means that the valuation must also be independent of the final outcome. It also means that where a chance has been lost, it is actionable regardless of whether or not the illness was cured. If a claimant suffers the physical harm and has a successful claim in respect of that outcome he should still be able to recover for the loss of a chance because that was valuable to him during his treatment – it is a distinct head of damage.

In contrast, the claims that have been brought on the basis of loss of a chance have sought to recover a proportion of the value of the physical injury. In *Hotson* the trial judge found that the claimant had lost a 25 percent chance of avoiding avascular necrosis and awarded 25 percent of the value of this physical damage. In *Gregg*, the claimant argued that his chance of survival had been reduced from 42 percent to 25 percent and the claim was for a proportion of the value of the loss that he would suffer if he failed to survive i.e. a proportion of the value of the physical injury itself and of the loss of lifetime's earnings.

Hogg has argued that proportionate recovery is a fair solution:

[L]ost chance recovery provides an equitable *via media* to the problem of causal uncertainty...Such an approach strikes a balance between letting off a defendant who may in fact have caused the harm, and penalising a defendant who may not have caused the harm at all.¹¹⁶

This justification is based on economic efficiency or on vague notions of fairness and is incompatible with corrective justice based liability. Voyiakis explains:

¹¹⁶ Martin Hogg, 'Re-establishing Orthodoxy in the Realm of Causation' (2007) Edin LR 8, 15.

Knowing how much risk I have imposed on you and how much compensation I would have to pay you if I had caused you actual physical harm does not by itself suggest how much I should pay for having exposed your physical health to danger. To take a crude example, when I make it 40 per cent more likely that you will suffer lung cancer, I am not causing you 40 per cent of the harm that lung cancer brings about. The quantity of risk and the quantity of physical harm do not seem to be related in any such straightforward way. We therefore have reason to object to the latter being taken as a measure of the former.¹¹⁷

This much is apparent if we recall the more straightforward causal models addressed in Chapter Two. There it was said that if the harm is indivisible then each of the necessary elements of the sufficient set is a cause and each is a cause of the whole loss. Apportionment of *liability* might be appropriate as a matter of moral responsibility, but each cause has causal responsibility for the whole of the harm. In contrast, where the harm is divisible i.e. dose-related, then each causal factor contributes/causes a portion of the overall harm. Thus a defendant who has contributed 40 per cent of the harmful agent has caused 40 per cent of the harm. So it seems illogical to say that a doctor who has contributed a 40 percent risk to an indivisible harm should pay 40 percent of the value of the physical harm. Valuing the chance as a percentage of the physical harm reinforces the sense that the loss of a chance argument as it was advanced in *Hotson* and in *Gregg* is simply an attempt to assist a claimant to bypass the causation requirement.

The argument made here, instead, is that the loss of a chance is an actionable loss independent of any physical harm. It should, therefore, be treated as any other loss and assigned a value within the Judicial College's *Guidelines for the Assessment of General Damages in Personal Injury Cases*.¹¹⁸ Jansen has addressed the 'pragmatic objection [which] emphasizes the difficulties of assessing the value of chances':

¹¹⁷ Emmanuel Voyiakis, 'The Great Illusion: Tort Law and Exposure to Danger of Physical Harm' (2009) 72 MLR 909, 917.

¹¹⁸ (n83).

[T]here are difficulties in assigning a financial value to lost chances, especially if they are related to physical harm...It is difficult to evaluate physical injuries financially. The widely differing amounts which courts award to physically injured parties illustrate that. But this indicates only that there is no ‘objective’ measure as such, and that the measure, therefore is to be determined by policy considerations. Courts do manage to deal with this consistently.¹¹⁹

In other words, the value assigned to any physical injury is necessarily artificial. This has been addressed by using the tariff system so that the level of damages is commensurable with the seriousness of the injury. Likewise, in relation to risk, Voyiakis argues that while the value of the risk is not necessarily an equivalent percentage of the physical harm, we have an intuitive notion that some risks are more serious than others. He explains:

This intuition seems to follow naturally from the idea that protecting oneself against risk of some harm is a means to protecting oneself against the actual harm. It also seems to tally with our intuitive comparisons across risks of different seriousness: the effect of being exposed to a 40 per cent risk of lung cancer is more serious than the effect of being exposed to a 40 per cent risk of a broken toe exactly because getting lung cancer is much more serious than getting a broken toe.¹²⁰

This means that it would be possible to set a tariff for a lost chance. Any physical injury is represented by a range of values in the tariff, with the judge being able to use the range of the scale to reflect the impact the injury had on the individual. ‘Loss of a chance’ could have a similar scale. It would be possible to set a value for the loss of a chance of avoiding a life-threatening illness with a scale of damages to reflect a small reduction in chance or a large reduction in chance, another value for the loss of the chance of avoiding a less serious illness (or disability or temporary illness) with a similar scale to reflect the extent of the reduction in the chance.

¹¹⁹ Jansen (n102) 293.

¹²⁰ Voyiakis (n117) 917.

Conclusion

In summary, the loss of chance claim for proportionate recovery is rejected in this thesis because a patient does not have an objective chance of being cured. This means that proportionate recovery effectively awards damages for the physical harm with a discount to reflect the degree of doubt surrounding proof of causation. Instead this thesis proposes a more limited role for loss of a chance as an independent form of damage leading to limited recovery. Such loss only arises in a meaningful way because of the special nature of the doctor-patient relationship, and only in relation to misdiagnosis/mistreatment of existing illness.

Chapter Four: The evidentiary gap

Introduction

This chapter now addresses the problem of the ‘evidentiary gap’ that arose in *McGhee*¹ and in *Fairchild*.² In this chapter it is necessary first to analyse the relevant case law in order to understand what exactly the so-called ‘evidentiary gap’ is. The first section will draw on the NESS analysis of causation presented in chapter two to locate and fully understand the problem that arose in these cases. Whereas claimants more commonly face the problem of proving that the negligence was a *necessary* element of a sufficient set of conditions that actually occurred, it will be shown that where there is an ‘evidentiary gap’ the claimant faces the prior problem that the state of scientific uncertainty prevents the definition of a sufficient set of conditions for the harm that he has incurred. The subsequent sections will turn to the legal solutions that have been adopted to address the evidentiary gap. Beginning with *McGhee*, and continuing through *Fairchild*, and *Barker*,³ the courts have considered it sufficient for the claimant to prove that the defendant’s negligence materially increased the risk of the relevant injury, or materially contributed to the risk of the relevant harm, in cases involving an evidentiary gap.

One interpretation of this test is that it simply alters the legal standard of proof that the claimant is required to meet, so section two will consider whether this is an appropriate solution. Related aspects of proof will also be considered, such as the place of statistics and the possibility of proof of causation on traditional principles through proof that the negligence more than doubled the background risk. Ultimately it will be argued that given the gap in scientific understanding surrounding the aetiology (the ‘evidentiary gap’) it is inappropriate for the law to attempt to bridge this gap – such resort to intuition in the face of explicit scientific uncertainty would base

¹ *McGhee v National Coal Board* [1973] 1 WLR 1 (HL (Sc.)).

² *Fairchild v Glenhaven Funeral Services* (t/a GH Dovener & Son) [2002] UKHL 22, [2003] 1 AC 32.

³ *Barker v Corus* [2006] UKHL 20, 2 WLR 1027.

liability on a fiction. The result is that the defendant is liable even though his negligence may not have been a cause of the harm.

Section three will therefore turn to the alternative interpretation of the *Fairchild* test, as proposed in *Barker*, that in cases where the test applies the damage which forms the gist of the negligence action is the increase in risk.⁴ Liability under *Barker* was several (rather than joint and several) and the extent of an individual defendant's liability was to be measured by the extent of his contribution to the total risk to which the claimant was exposed. At first glance this solution appears to conform to corrective justice because the defendant is held liable only for that loss which he has been proved to have caused i.e. since he has only been proved to have increased the risk of the claimant's injury, he is only held liable for this increase in risk. However, closer analysis shows a number of flaws in this approach. By insisting that the claimant have suffered the relevant physical harm before allowing him to bring a negligence action for exposure to the risk of this harm the effect is that the physical harm remains the gist of the action and the apportionment of liability that takes place reflects the possibility that the defendant caused the harm. (This echoes the criticism of proportionate recovery for 'loss of a chance' – the reality is that it is liability for possible causation of harm.) Furthermore, liability for risk is a more deeply problematic concept. Risk is a forward-looking concept so is mismatched with the backward-looking causation enquiry. More significantly, risk lacks moral significance as damage in the context of the negligence enquiry because it does not add anything to the breach of duty requirement. Conduct is wrongful if the defendant failed to take reasonable care i.e. he exposed the claimant to an unreasonable risk of harm. If 'exposure to risk' also constitutes actionable damage then the damage requirement would effectively be subsumed into the breach inquiry.

⁴ Although the majority in *BAI v Durham* [2012] UKSC 14, [2012] 1 WLR 867 considered that the gist of the action was still the physical harm and that the decision in *Barker* had simply equated 'materially increasing the risk' with 'contributing to the cause', Lord Phillips (dissenting) insisted that the rule in *Barker* was that where the *Fairchild* test of 'material increase in the risk of harm' is applied the gist of the action is the risk of harm.

This is inconsistent with corrective justice which is concerned with ‘wrongful loss’ not with wrongdoing in isolation.⁵

The final section will refocus on corrective justice and argue that these claims ought to fail for want of proof of causation since the approaches explored in previous sections fail to achieve consistency with corrective justice. Since the defendant is liable in circumstances where he may not have been a cause of the loss, corrective justice is abandoned in these cases in favour of the pursuit of consequentialist goals. The effect of this has been to throw negligence into the state of incoherence that is inherent in the pursuit of consequentialist goals within the bipolar framework of the tort action. *Barker* seems to achieve a fair balance between the competing interests of claimants and defendants, but this is simply the author’s personal preference since there is no logical demand on the limits of liability. Ultimately it is preferable that the scope of application of the *Fairchild* test be as tightly circumscribed as possible so that the incoherence that has affected liability in these cases is not repeated in other contexts.

1. Defining the evidentiary gap

The so-called ‘evidentiary gap’ problem was first addressed in *McGhee v National Coal Board*⁶ in relation to the cause of dermatitis, and subsequently in relation to the cause of mesothelioma in *Fairchild v Glenhaven Funeral Services* (and later cases).⁷ The defendant in *McGhee* was held to be liable for the claimant’s dermatitis on the basis that his negligence had materially increased the risk of dermatitis. Following this decision it was unclear whether this constituted an exceptional new ‘test’ for causation or whether the court had simply taken a ‘robust’ view of the evidence

⁵ Porat and Stein have also suggested that the gist of the action in evidentiary gap cases could be reconceptualised as ‘evidentiary damage’. They suggest that if there was only one source of the harmful agent the claimant would not face any problem of proof, so each defendant who exposes the claimant to another source of the harmful agent is responsible for preventing the claimant from having a successful negligence action. This argument has not been widely accepted, and raises a distinct question as to whether a claimant has a right to protect the integrity of his evidence from indirect interference. The evaluation of this right falls outside the scope of this thesis. See Ariel Porat and Alex Stein, *Tort Liability Under Uncertainty* (Oxford University Press 2001).

⁶ *McGhee v National Coal Board* 1972 SLT (notes) 61 (Court of Session, Scotland); [1973] 1 WLR 1 (HL (Sc.)).

⁷ *Fairchild v Glenhaven Funeral Services* (t/a GH Dovener & Son) [2001] EWCA Civ 1881, [2002] 1 WLR 1052; [2002] UKHL 22, [2003] 1 AC 32.

before it. The House of Lords held that the test did not apply to the facts of *Wilsher v Essex AHA*,⁸ but did apply in *Fairbeld*. The purpose of this section is to examine the causal problem that arose in each of these cases, and the scientific evidence that was available, in order to understand precisely what the ‘evidentiary gap’ consists of. Approaching this from the basis of the NESS test will be shown to have a dual advantage over the but-for test. It first enables the causal problem to be conceptualised more accurately. This then means that the possible legal responses to the problem of the evidentiary gap can be analysed with greater clarity.

1.1 The evidentiary gap relating to dermatitis: *McGhee v National Coal Board*

The claimant in *McGhee* developed dermatitis as a result of damage to his skin from brick dust. He became covered in dust while emptying brick kilns for the defendant employer. This exposure to brick dust at work was unavoidable, so it was not negligent. The defendant’s negligence consisted of failure to provide shower facilities to enable the claimant to wash the brick dust off his skin before cycling home. The medical evidence was that the claimant’s dermatitis was caused by brick dust on his skin, so both the innocent and guilty periods of exposure to brick dust were potential causes of the dermatitis in this case. The problem facing the claimant was to prove that the negligent exposure in particular was an actual cause of his dermatitis. On the basis of the available legal tests he was required to prove that but-for the negligent exposure he would not have developed dermatitis, or, applying the *Wardlaw* test,⁹ that the negligent exposure had materially contributed to his dermatitis.

The difficulty the claimant faced in this particular case was that although medical science was able to state that brick dust *does* cause dermatitis, it was unable to explain *how* it causes dermatitis.¹⁰ This meant that proof that the negligent exposure had actually contributed to the claimant’s disease was problematic. Lord Wilberforce noted that the experts ‘had to admit that

⁸ *Wilsher v Essex AHA* [1988] AC 1074 (HL).

⁹ *Bonnington Castings v Wardlaw* [1956] AC 613 (HL).

¹⁰ *McGhee v National Coal Board* [1973] 1 WLR 1 (HL) 3 (Lord Reid).

they knew little of the quantity of dust or the time of exposure necessary to cause a critical change'.¹¹ Lord Reid explained this gap in the medical experts' understanding more clearly:

In the present case the evidence does not show – perhaps no one knows – just how dermatitis of this type begins. It suggests to me that there are two possible ways. It may be that an accumulation of minor abrasions of the horny layer of the skin is a necessary precondition for the onset of the disease. Or it may be that the disease starts at one particular abrasion and then spreads, so that multiplication of abrasions merely increases the number of places where the disease can start and in that way increases the risk of its occurrence.¹²

The 'evidentiary gap' in this case therefore concerns the aetiology of the disease, in particular the question of whether it is caused by an accumulation of abrasions or by a single abrasion. If it were caused by a single abrasion then the but-for test would apply, and if it were caused by an accumulation of abrasions the test of 'material contribution to harm' would apply. Since the causal process is unknown, neither test can be satisfied.

1.2 The evidentiary gap relating to mesothelioma: *Fairchild v Glenhaven Funeral Services*

The victim in *Fairchild* had died from mesothelioma, a malignant tumour in the pleura (the membrane surrounding the lungs) and the negligence action was brought by his widow.¹³ During his working life he had been exposed to asbestos by a number of his former employers, one of whom was the defendant employer. The problem facing the claimant was to establish that the defendant's negligence was a cause of the victim's disease. Similarly to *McGhee* it was known that inhalation of asbestos fibres *does* cause mesothelioma, but medical science was unable to explain *how* it causes mesothelioma. Once again, therefore, the problem of proof arose because of the

¹¹ *ibid* 5 (Lord Wilberforce).

¹² *ibid* 4 (Lord Reid).

¹³ *Fairchild v Glenhaven Funeral Services Ltd and others* [2002] UKHL 22, [2003] 1 AC 32, [7].

scientific uncertainty combined with the fact that there was more than one period of exposure to the harmful substance.

Lord Bingham explained the uncertainty:

The mechanism by which a normal mesothelial cell is transformed into a mesothelioma cell is not known. It is believed by the best medical opinion to involve a multi-stage process, in which six or seven genetic changes occur in a normal cell to render it malignant. Asbestos acts in at least one of those stages and may (but this is uncertain) act in more than one...It is accepted that the risk of developing a mesothelioma increases in proportion to the quantity of asbestos dust and fibres inhaled: the greater the quantity of dust and fibre inhaled, the greater the risk. But the condition may be caused by a single fibre, or a few fibres, or many fibres: medical opinion holds none of these possibilities to be more probable than any other, and the condition once caused is not aggravated by further exposure.¹⁴

The scientific uncertainty concerning the quantity of asbestos fibres required to cause mesothelioma, and the stage(s) at which the fibres are involved in the causal process, combined with the fact that the claimant had been exposed to asbestos by a number of former employers, meant that he was unable to establish causation, either on the but-for test or on the ‘material contribution to harm’ test because it was impossible to say what he needed to prove. Stapleton has explained:

This means that there is as yet no means of telling whether the mesothelioma of a person subjected to a sequence of asbestos exposures was due to all exposures, only some or one, let alone which one. Importantly, there is no direct basis for saying that longer, more intense exposures are more likely to have been the cause...than much shorter

¹⁴ ibid (Lord Bingham).

exposures, nor is there any basis for saying that earlier exposures are more likely to have been the cause than later exposures.¹⁵

Lord Bingham called this a ‘rock of uncertainty’, and it is generally known as the ‘evidentiary gap’.¹⁶ The evidentiary gap is therefore a very specific evidential problem. It is not simply a problem of a lack of evidence relating to the individual claimant. It goes to the general state of knowledge and understanding of the disease in medical science. The ‘evidentiary gap’ must also be described more accurately than ‘scientific uncertainty about the aetiology of the disease’. The *disease* is known to be indivisible (i.e. its severity is not dose-related), and the harmful agent is known (e.g. brick dust, asbestos), but the *process* by which it is caused may be an accumulation or interaction of a few or many fibres.¹⁷

Little has changed in the understanding of the aetiology of mesothelioma in the years since *Fairchild*, as Lord Phillips explained in the recent case of *Sienkiewicz v Greif*.¹⁸ It is still believed that a small number of cases may possibly be idiopathic, meaning they have an unknown cause other than asbestos, but mesothelioma is ‘always, or almost always, caused by the inhalation of asbestos fibres’.¹⁹ One development since the decision in *Fairchild* concerns the minimum length of the latency period which is now believed to be five years rather than ten. This period is important because any exposure to asbestos during this period has no causative effect because the disease has already been caused. Furthermore, it is no longer believed that mesothelioma is

¹⁵ Jane Stapleton ‘Lords a’leaping evidentiary gaps’ (2002) 10 Torts LJ 276, 281.

¹⁶ See for example Stapleton ‘Lords a’leaping’ (n15).

¹⁷ It is therefore distinct from the problem that arose in *Wilsher*. The claimant there was a baby born prematurely who developed retrosternalfibroplasia (‘RLF’). The defendant had negligently exposed the claimant to excess oxygen which is known to cause RLF, but the claimant also suffered a number of other conditions that were equally capable of causing RLF and which were due to his premature birth rather than to negligence. The case will be examined in greater detail at a later stage, but it is important to note that the problem in that case was not that medical science was unable to explain how RLF is caused, but that the claimant was unable to show that excess oxygen was a cause in his individual case.

¹⁸ *Sienkiewicz v Greif (UK) Ltd* [2011] UKSC 10, [2011] 2 AC 229. Lord Phillips based this understanding on a case control study by Peto and Rake, published in 2009 by the Health and Safety Executive, on “Occupational, Domestic and Environmental Mesothelioma risks in Britain” (“the Peto Report”), which he explained ‘is said to be the first representative study to quantify the relationship between mesothelioma risk and lifetime occupational and residential history in this country’ (at [18]), and on ‘the collation of data about mesothelioma set out by Rix LJ in his judgment in the series of appeals collectively described as *Employers’ Liability Insurance ‘Trigger’ Litigation* [2010] EWCA Civ 1096’ (at [19]).

¹⁹ *Sienkiewicz* (n18) [19] (Lord Phillips).

triggered by a single asbestos fibre.²⁰ While it is therefore believed that mesothelioma is caused by multiple asbestos fibres, there remains significant scientific uncertainty as to how it is caused and at what stages the fibres are effective. Lord Phillips explained:

It is believed that a cell has to go through 6 or 7 genetic mutations before it becomes malignant, and asbestos fibres may have causative effect on each of these.

It is also possible that asbestos fibres have a causative effect by inhibiting the activity of natural killer cells that would otherwise destroy a mutating cell before it reaches the stage of becoming malignant.²¹

He also noted another possible causal mechanism, explaining that ‘The Peto Report also raised the possibility (but no more) of synergistic interaction between early and later exposures. Causation may involve a cumulative effect with later exposure contributing to causation initiated by an earlier exposure’.²²

1.3 Analysis of the evidentiary gap

The focus of the chapter has so far been confined to determining the nature of the evidentiary gap. The solution that was adopted in *McGhee* and applied in *Fairchild* was to hold the defendant liable on the basis that his negligence had materially increased the risk of the harm suffered. Before it is possible to engage in an analysis of this solution, and of the subsequent developments to it in later cases, it is essential to understand more precisely the causal problem that arises in these cases.

Steel and Ibbetson have explained that, ‘[t]he normal rule of causation...has two aspects: the evidential and the conceptual. In order to establish liability, P has to show on balance of probabilities (evidential) that but for D’s wrongful conduct the injury would not have occurred

²⁰ ibid [102] (Lord Phillips).

²¹ ibid [19] (Lord Phillips).

²² ibid [102] (Lord Phillips).

(conceptual).²³ Problems of causation divide into these two categories, the conceptual and the evidential. Stapleton has explained that the problem in evidentiary gap cases is an evidential one: '[the claimant's] difficulty was the antecedent one that there was simply inadequate factual information to feed into the but-for test'.²⁴ She therefore considers it irrelevant to address the conceptual inadequacies of the but-for test because 'this was not the problem facing the claimants in *Fairchild*'.²⁵ In this section it will be argued, however, that there are a number of advantages to adopting the NESS test which was shown in chapter two to be conceptually more robust than the but-for test. First, since the NESS test is better matched to the underlying concept of causation it enables us to pinpoint precisely which aspect of causation the evidential problem relates to. Second, the NESS analysis highlights the fact that the *Wardlaw* test of material contribution to harm is distinct in nature from the *McGhee/Fairchild* test of material contribution to the risk of harm. The former addresses a conceptual weakness in the but-for test, so does not depart from the causation requirement, the latter addresses an evidential obstacle to proof of causation regardless of how causation is conceptualised so represents a significant exception to the causation requirement. Adopting the NESS test would not solve the evidential problem raised by the evidentiary gap, but it would make it easier to understand exactly what is at stake in these cases.

1.3.1 NESS and identifying the problem: locating the evidentiary gap

In Wright's NESS account of causation, 'a particular condition was a cause of (condition contributing to) a specific consequence if and only if it was a necessary element of a set of antecedent actual conditions that was sufficient for the occurrence of the consequence'.²⁶

The problem that a claimant usually faces in negligence is to establish the *necessity* of the defendant's negligence in order to complete the sufficient set of conditions. An example is the

²³ Sandy Steel and David Ibbetson, 'More grief on uncertain causation in tort' (2011) 70 CLJ 451, 452.

²⁴ Stapleton, 'Lords a'leaping' (n15) 280.

²⁵ *ibid.*

²⁶ Richard Wright, 'Causation in Tort Law' (1985) 73 Cal L Rev 1735, 1790.

so-called ‘single hit hunters’ scenario that arose in the Canadian case of *Cook v Lewis*.²⁷ If two hunters have each negligently fired their guns in the victim’s direction and the victim has been hit by only one bullet then the problem he faces is proving which hunter fired the bullet that hit him i.e. which bullet was *necessary* for the sufficient set of conditions? Additionally the cases of *Hotson v East Berkshire HA* and *Gregg v Scott* that were examined in the previous chapter involved an evidential problem concerning the necessity of the negligence, and the claimants attempted to avoid this problem of proof by proposing ‘loss of a chance’ as an alternative form of damage. It was clear in these cases what they were required to prove, the difficulty was in proving it to the balance of probabilities standard of proof. For example, in *Hotson* the sufficient set of conditions for avascular necrosis is ‘insufficient intact blood vessels to keep the epiphysis alive’. The problem for the claimant was to prove whether those blood vessels that were damaged during the delay caused by the defendant’s negligence were *necessary* for this set, or whether they had been pre-empted from having any effect by the quantity of blood vessels damaged during the fall.²⁸

In contrast, the claimants in the evidentiary gap cases of *McGhee* and *Fairchild* faced the prior problem of identifying what constitutes a *sufficient set* of conditions to cause the relevant disease. This was essential to their claims because there were multiple exposures to the harmful agent so there were multiple potential causes. If medical science is unable to explain what constitutes a sufficient set of conditions, then a claimant faces an insurmountable challenge – how can he prove that the defendant’s negligence was necessary for the sufficiency of the set of conditions to cause his dermatitis or mesothelioma if uncertainty surrounding the aetiology of his disease means that medical science is unable to tell him what a sufficient set consists of? In other words, in most negligence cases we know what the claimant needs to prove and the problem for him is proving it to the requisite legal standard of proof. In cases involving this kind of evidentiary gap

²⁷ [1951] SCR 830

²⁸ See Ch 3 section 3.1.2.

it is not possible to say what the claimant needs to prove let alone asking him to prove it to a particular standard.

The NESS test therefore enables the distinction to be made clearly between evidence of causation in the personal sense of proving that the negligence was *necessary* for the sufficiency of the set of conditions, and evidence of causation in the deeper, general sense of proving what a sufficient set of conditions is for a given disease. The next section builds on this NESS-informed understanding of the evidentiary gap in order to explore possible legal solutions with greater clarity.

1.3.2 NESS: separating the conceptual from the evidential aspects of the legal solution

The solution adopted by the majority of the House of Lords in *McGhee* built on the *Wardlaw* test of material contribution to harm and suggested that there is ‘no substantial difference between saying that what the defendant did materially increased the risk of injury to the pursuer and saying that what the defendant did made a material contribution to the injury’.²⁹ It will be argued that there is, however, a significant difference between these two propositions since they are different in nature. The *Wardlaw* test addresses a conceptual inadequacy of the but-for test whereas *McGhee* addresses the evidentiary gap so it assists a claimant who faces an evidential barrier to proof on any test. Therefore, *McGhee* is not simply an extension of the *Wardlaw* test, it is an exception to it.

The problem with the decision in *McGhee* is that it seems to be built upon the misconceived idea that the *Wardlaw* test of ‘material contribution to harm’ is an exceptional approach which facilitates proof of causation and that making a further allowance where the claimant can only prove a material increase in the risk of harm is not a significant leap. This is partly due to the fact that, as chapter two illustrated, the speeches in *Wardlaw* did not adequately define the causal or

²⁹ *McGhee* (n10) 5 (Lord Reid).

evidential problems in that case so the scope and effects of the test of material contribution to harm are unclear. But it is also partly because the analysis in *McGhee* failed to define accurately the causal and evidential problems raised by either case.

In *McGhee*, Lord Salmon and Lord Reid both observed that a distinction could be drawn between the causal processes in *Wardlaw* and *McGhee* yet they chose to dismiss it as irrelevant. Lord Reid explained that dermatitis may be caused by an accumulation of abrasions, or by one particular abrasion, but continued to say:

I am inclined to think that the evidence points to the former view. But in a field where so little appears to be known with certainty I could not say that that is proved. If it were, then this case would be indistinguishable from *Wardlaw's* case.³⁰

He explained:

Nor can I accept the distinction...between materially increasing the risk that the disease will occur and making a material contribution to its occurrence.

There may be some logical ground for such a distinction where our knowledge of all the material factors is complete. But it has often been said that the legal concept of causation is not based on logic or philosophy. It is based on the practical way in which the ordinary man's mind works in the everyday affairs of life. From a broad and practical viewpoint I can see no substantial difference between saying that what the defendant did materially increased the risk of injury to the pursuer and saying that what the defendant did made a material contribution to his injury.³¹

This highlights that the ordinary man's analogy between increasing the risk of harm and contributing to the harm is flawed because he had previously explicitly stated that causation of dermatitis is *not* analogous with the pneumoconiosis in *Wardlaw*. The danger of regarding *Wardlaw* as taking an exceptional approach is that courts will not analyse the conceptual and evidential problems that necessitated the 'material contribution to harm' approach, and will

³⁰ *ibid* 4 (Lord Reid).

³¹ *ibid* 5 (Lord Reid). See also 12-13 (Lord Salmon).

instead focus on contextual similarities and policy considerations. Although Lord Wilberforce adopted a different solution in *McGhee*, reversing the onus of proof, he still drew on contextual similarities to *Wardlaw* in order to justify taking an exceptional approach. He focused on the fact that both cases involved industrial disease and explained ‘[i]n cases concerned with pneumoconiosis, the courts faced with a similar, though not identical, evidential gap, have bridged it by having regard to the risk situation of the pursuer’.³² Likewise, in the later decision in *Fairchild*, Lord Nicholls held that ‘[o]n occasions the threshold “but for” test of causal connection may be over-exclusionary. Where justice so requires, the threshold itself may be lowered’.³³

It is essential to understand why the but-for test is over-exclusionary in order to assess whether justice requires an exceptional approach. In the case of *Wardlaw*, the exception to the but-for test was necessary because the but-for test was conceptually inadequate as a test of causation. In *McGhee* and *Fairchild* the problem is not the conceptual aspect of the test for causation, but the deeper evidential problem that it is not possible to define a sufficient set. As already stated, adopting the NESS test would not overcome the problems facing the claimant because the problem is an evidential one. The benefit of the NESS test is that it shows us that it is only possible to recognise liability in these cases if an exceptional approach is adopted.

Now that the problem of the evidentiary gap has been explained, the focus of the chapter can turn to the question of how to address this problem in a way that is consistent with corrective justice. Sections Two and Three will consider a range of legal and academic responses in terms of the concepts involved, before section Four returns the focus squarely on the demands of corrective justice. Solutions fall into two broad categories: those that focus on the legal standard of proof and those that seek to avoid the problem by reformulating the damage that forms the gist of the negligence action. Section two is therefore focused on proposals that relate to the

³² *ibid* 6 (Lord Wilberforce).

³³ *Fairchild* (n13) [40] (Lord Nicholls).

evidential standard, and section three considers attempts to reformulate the damage in these cases.

2. Evidential solutions: the inferential interpretation of *McGhee*

The majority in *McGhee* found that the defendant had materially increased the risk of dermatitis, and considered this a sufficient basis for liability. Lord Kilbrandon did not resort to reasoning based on a material increase in risk, simply finding that causation had been established on the balance of probabilities. Lord Wilberforce, however, pointed to the contradiction that lay at the heart of the majority approach. He had made a similar assessment of the facts of the case, explaining that pneumoconiosis involves a ‘similar, though not identical, evidential gap’.³⁴ Yet he preferred to resolve the problem by reversing the onus of proof because ‘to bridge the evidential gap by inference seems to me something of a fiction, since it was precisely this inference which the medical expert declined to make’.³⁵

There were two competing interpretations of the significance of the finding of a ‘material increase in risk’ following *McGhee*. It was initially held in *Wilsher* that the majority had simply taken a ‘robust and pragmatic’ view of the available evidence in order to draw in inference of causation.³⁶ It was eventually held in *Fairchild* that this was an exceptional test for causation that was applicable in tightly circumscribed cases to enable claimants to overcome the evidentiary gap.³⁷ However, in his corrective justice-based account of the tort of negligence, *Rediscovering the Law of Negligence*, Beever argues that the majority in *McGhee* did draw a factual inference of causation and that they were entitled to find that the claimant had succeeded in proving the causal link on the balance of probabilities.³⁸ Since the main arguments about causation in this thesis are premised on a corrective justice account of negligence it is essential to examine the validity of Beever’s assertions.

³⁴ *ibid* 6 (Lord Wilberforce).

³⁵ *ibid* 7 (Lord Wilberforce).

³⁶ *Wilsher* (n8) 1090 (Lord Bridge).

³⁷ *Fairchild* (n13) [22] (Lord Bingham), [44] (Lord Nicholls), [65] (Lord Hoffmann), [142] (Lord Rodger).

³⁸ Allan Beever, *Rediscovering the Law of Negligence* (Hart Publishing 2007) 465-472.

It will be argued that the nature of the evidentiary gap means that it is simply not possible to establish on the balance of probabilities that the defendant's negligence was a cause of the claimant's disease. It may be the case that the scientific community generally requires a higher standard of proof than the legal standard of the balance of probabilities, so there will be cases where the expert evidence does not prove causation definitively yet a court can still justifiably conclude that causation has been established on the balance of probabilities. But where the expert evidence demonstrates an evidentiary gap there is simply insufficient information to feed in to any test or in to the balance of probabilities standard of proof – no explanation is more likely than any other. For a court to find that causation has been established on the balance of probabilities when the scientific evidence is that causation *cannot* be established because of an evidentiary gap, is to make a deliberately misinformed judgment based on intuition both as to causation and as to overall responsibility for the loss. It is one thing for the court to be satisfied on the balance of probabilities where there is some degree of doubt or uncertainty in the expert evidence because the law requires only probability not certainty, but it is entirely different for a court to resort to intuition about what probably happened where the expert evidence clearly states that causation cannot be established.

2.1. The inferential approach to *McGhee*

As Lord Reid explained, the medical evidence in *McGhee* showed that brick dust can cause dermatitis, and that the longer it is on the skin the greater the risk of developing dermatitis, but could not explain why this is so. He conjectured that:

Plainly that must be because what happens while the man remains unwashed can have a causative effect, although just how the cause operates is uncertain. I cannot accept the view...that once the man left the brick kiln he left behind the causes which made him

liable to develop dermatitis. That seems to me quite inconsistent with a proper interpretation of the medical evidence.³⁹

According to Beever ‘the most natural reading’ of this passage is that ‘Lord Reid said that, on the balance of probabilities, the defendant’s negligence caused the claimant’s injury’.⁴⁰ He says that this view was also supported by Lords Simon, Kilbrandon and Salmon so that they all ‘rightly or wrongly, insisted that the claimant in *McGhee* could establish causation on the balance of probabilities’.⁴¹ It is clear that there are two issues at stake here, first whether Lord Reid *did* find causation to be established on the balance of probabilities rather than applying a new principle of law and, second, whether the evidence in the case was actually capable of supporting such a finding.

The first of these questions was resolved by the House of Lords in *Fairchild* where the majority held that the facts of *McGhee* simply did not support a finding of causation on the balance of probabilities and, given that Lord Reid had explicitly noted the distinction between the facts of *McGhee* and *Wardlaw*, he cannot have intended to make a logically flawed finding of fact. This means that the more interesting and more important question is the second question – was the evidence in *McGhee* capable of supporting a finding of causation on the balance of probabilities?

Beever says that where there is scientific uncertainty ‘[c]rucially, the Court cannot simply side with the defendant when the expert witnesses cannot or will not express their views on probabilities’.⁴² There is some truth to this assertion but it needs to be heavily qualified. The burden of proof in civil law rests on the claimant, so if the claimant is unable to persuade the court of his claim on the balance of probabilities then it must fail. If the causal process is understood i.e. it is known what constitutes a sufficient set, and the question is whether the defendant’s negligence was necessary for the sufficiency of that set, then some pragmatism is needed in law and care needs to be taken to see what the scientific evidence tells us and what are

³⁹ *McGhee* (n10) 4-5 (Lord Reid), cited by Beever (n38) 466 (emphasis Beever’s).

⁴⁰ Beever (n38) 466.

⁴¹ *ibid*.

⁴² *ibid* 471.

the limits of what it does and doesn't say. But in cases involving an evidentiary gap problem the scientific evidence is that it is impossible to say what constitutes a sufficient set of conditions. This means that the experts' refusal to say how likely it is that the negligence was necessary for the sufficiency of the set is not a symptom of their adherence to a higher standard of proof than the legal standard, but instead is based on the scientific impossibility of assigning a probability. In other words, where it is theoretically possible to determine whether the negligence was a cause but the expert witnesses are reluctant to express an opinion because they are observing a higher standard of proof then it is still appropriate for the judge to assess his own degree of belief, but where the evidentiary gap makes it impossible for the expert witnesses to express an opinion there is no logical basis for the judge to have any degree of belief.

2.1.1 Scientific and legal standards of proof

Beever says that the medical experts in *McGhee* 'said that it was not possible to prove...that the defendant's negligence caused the claimant's injury, but, as the medical experts were not applying the balance of probabilities, that does not imply that it was proved that the defendant did not cause the plaintiff's injury'.⁴³ His argument therefore centres on the different standards of proof in tort law and in science.

Courts have noted that standards of proof are higher in science than the civil law's balance of probabilities standard. Lord Prosser in *Dingley v The Chief Constable, Strathclyde Police* explained that:

Whether one uses the word 'scientific' or not, no hypothesis or proposition would be seen as 'proved' or 'established' by anyone with any form of medical expertise merely upon the basis that he had come to regard it as probably sound...And even if, in relation to any possible proposition or hypothesis, such an expert even troubled to notice that he had come to the point of regarding it as not merely possible but on balance 'probable', then I think he would regard that point as one from which he must set off on further

⁴³ ibid 467.

inquiry, and by no means as being (as it is in the courts) a point of arrival. Mere marginal probability will not much interest him. But it must satisfy a court.⁴⁴

Lord Phillips also noted the difference between legal and scientific standards in the later case of *Sienkiewicz v Greif (UK) Ltd*:

When a scientific expert gives an opinion on causation, he is likely to do so in terms of certainty or uncertainty, rather than probability. Either medical science will enable him to postulate with confidence the chain of events that occurred, i.e. the biological cause, or it will not. In the latter case he is unlikely to be of much assistance to the judge who seeks to ascertain what occurred on a balance of probability.⁴⁵

In the context of a negligence claim there is a clear place for pragmatism in the law. The courts cannot postpone decisions until scientific certainty is achieved. As Lord Rodger said in *Fairchild* ‘the House must deal with these appeals on the basis of the evidence as to medical knowledge today and leave the problems of the future to be resolved in the future’.⁴⁶ Similarly, Miller has highlighted that ‘civil courts cannot commission laboratory experiments or epidemiological studies, but nor can they suspend a case until someone else does’.⁴⁷ This echoes the argument made in Chapter Two that whilst the legal test of causation ought to reflect a philosophically and scientifically correct account of causation so far as is practicable, the law is also shaped by the need to produce concrete decisions without delay.

But scientific or medical expert evidence cannot be written off as too demanding. Instead it is essential that the courts engage with the evidence so that they understand what it does and does not tell them in order to assess how persuasive the evidence is. This goes beyond identifying the statistics to which the expert witnesses refer, and requires engagement on a qualitative level also.

As Laleng has suggested:

⁴⁴ 1998 SC 548, 603.

⁴⁵ *Sienkiewicz* (n18) [9] (Lord Phillips).

⁴⁶ *Fairchild* (n13) [124] (Lord Rodger).

⁴⁷ Chris Miller, ‘Coal Dust, Causation and Common Sense’ (2000) 63 MLR 763, 769

an understanding of the similarities and differences between the two methods [legal and scientific] is more likely to yield acceptable results than policy-based decisions that seem to be based on little more than sympathy for victims and a possible generalised chemo-phobia.⁴⁸

The previous chapter explained at length that the balance of probabilities standard of proof is a ‘belief probability’.⁴⁹ The judge must be more convinced than not that the claimant’s account is true. A belief probability involves a qualitative assessment of the evidence, and in order to form a rational belief in causation in an individual case the court must assess not only the quantitative evidence and ‘fact probability’, but also the sampling error probability, the reliability of the studies from which the evidence is derived, and crucially must also attempt to personalise this to the individual claimant’s case.

The previous chapter drew heavily on Barnes’ article which addresses the relationship between legal and scientific standards of proof and seeks to dispel a number of misconceptions. He explains:

Judges and lawyers first encountering statistical evidence want to believe that scientific standards are tougher than legal standards. A court will reject an assumption that there is no causal connection between an act and an injury if the evidence makes causation “more likely than not.” A scientist will reject an assumption that there is no relationship between two variables only if there is less than a five percent probability that the statistical evidence showing a relationship is due to chance. The law appears willing to accept no more than a forty-nine percent chance of error while science appears willing to accept no more than a five percent chance of error.⁵⁰

⁴⁸ Per Laleng, ‘*Sienkiewicz v Greif (UK) Ltd and Willmore v Knowsley Metropolitan Borough Council: A Material Contribution to Uncertainty?*’ (2011) 74 MLR 777, 793.

⁴⁹ See Ch 3 section 3.2.

⁵⁰ David W Barnes, ‘Too Many Probabilities: Statistical Evidence of Tort Causation’ (2001) 64 Law and Contemp. Probs 191, 191.

He argues that '[t]his perception is incorrect, but hard to change'.⁵¹ As explained in chapter three, the five percent chance of error in science concerns the sampling error probability (p-value), and is not a belief probability so direct comparison is not possible. Furthermore, since a sampling error probability reflects only the chance that the subjects of a study happened to be atypical it can be calculated for a poorly-designed study as easily as for a well-designed study, so it would be a doing a disservice to scientific experts to suggest that they are only concerned with the p-value.

2.1.1.1 The place of intuition in the balance of probabilities

In *McGhee* Lord Kilbrandon said that proof of causation ‘depends on drawing a distinction between the possibility and the probability of the efficacy of the precautions. I do not find it easy to say in the abstract where one shades into the other’.⁵² The line between possibility and probability in the degree of belief is not a bright one because it necessarily involves a qualitative assessment of the evidence which introduces an element of subjectivity. But it is the role of the judge to decide on which side of the line a particular case falls. The qualitative aspect of the standard of proof does not, however, mean that the balance of probabilities standard of proof can be divorced from scientific evidence.

Stauch has criticised the balance of probabilities as being a crude test in comparison to statistical evidence:

[W]hereas statistics derived systematically from our previous experience of similar cases, provide us with a very accurate probability-weighting for each candidate [cause], the balance of probabilities test attempts to perform the same operation by appealing crudely to what we *feel* the likely cause to have been. The relevant feeling must, once again, derive from our previous experience of similar cases, but this time in its rawest form.⁵³

⁵¹ ibid.

⁵² *McGhee* (n10) 10 (Lord Kilbrandon).

⁵³ Marc Stauch, ‘Loss of Chance in Medical Negligence’ (1997) 17 OJLS 205, 219.

There are a number of important observations to make in relation to this statement. The first is that the balance of probabilities will necessarily seem less objectively measureable than statistical probabilities because it is a qualitative concept. It is a belief probability so it measures the judge's degree of belief in a proposition so necessarily involves his assessment of how credible the assertion is. Rather than being a weakness of the balance of probabilities standard of proof, Barnes work has shown that the qualitative aspect is vital to assessing the credibility of the scientific evidence and its value in an individual case. His work showed that in addition to the fact probability and sampling error probability the court must weigh up other factors affecting the reliability of the scientific study from which those probabilities were derived, and must extrapolate to the individual case. While statistical probabilities may have an outward appearance of objectivity, they only tell part of the story and they certainly do not provide a definitive answer about causation in an individual case.

Additionally, Stauch presents us with a choice between statistical probabilities and the balance of probabilities test and implies that by insisting on the balance of probabilities test the courts are making a conscious and deliberate decision to rely on a feeling rather than on scientific and statistical evidence. It is essential to remember that these two types of probability are not mutually exclusive, indeed, as Barnes' article showed, a rational belief probability must be based on an assessment of the credibility of the scientific evidence. It is important for judges to understand that the balance of probabilities standard requires them to form a belief based on the available evidence, rather than ignoring the scientific evidence in favour of a 'common sense' judgment of causation.

2.1.1.2 Evidence of the evidentiary gap

As previously explained, it is essential to personalise the evidence as far as possible to the individual claimant in order to form a rational belief about the causes of his injury and the

inferential approach loses sight of this vital step. The evidentiary gap means that it is not possible to define a sufficient set, and this prevents meaningful personalisation of the evidence.

Adopting the inferential approach, Beever argues that ‘[c]rucially, the experts did not say that there was no proof of a causal link between dust and dermatitis’.⁵⁴ This is true, but it was not suggested that the claimant was unable to prove the general link between brick dust and dermatitis. The problem was proving the link between the *negligent* exposure to brick dust and the dermatitis. Indeed, the fact that his dermatitis was caused by brick dust was what enabled the court to say that the negligent exposure to brick dust had increased the risk of suffering dermatitis. It is, however, a significant leap from saying that brick dust can cause dermatitis to saying that the negligent dust was a cause of this claimant’s dermatitis.

As Mullany clearly explains, for a belief in the fact of causation to be rational it must be based on scientific evidence since, ‘[t]he educative effect of the expert evidence makes appeals to common sense notions of causation largely meaningless or produces findings which would not be made by an ordinary person who had not been so instructed’.⁵⁵ In other words, if the ordinary person looked at the events in *McGhee* without the assistance of expert evidence then they would probably say that since brick dust is capable of causing dermatitis it makes sense to say that both periods of exposure to brick dust caused this claimant’s dermatitis. But once they were shown the medical evidence which said that dermatitis might be caused by an accumulation of abrasions, or it might just start at a single abrasion, the ordinary person would not be able to say that the negligent exposure contributed to the illness because they would see that it is as likely that it was caused by a single abrasion during the innocent exposure.

The NESS analysis showed that the evidentiary gap surrounding the aetiology of the disease means that it is not possible to define a ‘sufficient set’ of conditions. In cases such as *McGhee* and *Fairchild* where there are a number of sources of the harmful agent and an evidentiary gap preventing the definition of a ‘sufficient set’, it is simply not possible to form a rational belief

⁵⁴ Beever (n38) 468.

⁵⁵ Nicholas J Mullany, ‘Common Sense Causation – an Australian View’ (1992) 12 OJLS 431, 437.

that the negligence was necessary for the sufficiency of an actually occurring set of conditions when it is not possible to say what a sufficient set of conditions consists of. No matter how far the statistical and epidemiological evidence can be personalised to account for the individual's characteristics, and his working and living conditions, the evidentiary gap means that where there is more than one source of the harmful agent it is not possible to form a rational belief that a particular source was a cause. As the discussion of Barnes' work showed, for a belief probability to be rational it must be based on the available evidence and the evidence was clear that it was impossible to know because a sufficient set could not be defined.

Although the majority in *Fairchild* rejected the inferential approach, Lord Hutton adopted it, explaining that 'I consider that this approach, whereby the layman applying broad common sense draws an inference which the doctors as scientific witnesses are not prepared to draw, is one which is permissible'.⁵⁶ This is unjustifiable given that the problem was not that the scientific evidence was inconclusive, but that it showed a clear evidentiary gap in being unable to define a 'sufficient set'. This made it impossible to have a rational belief that the defendant's negligence was necessary for an actually occurring sufficient set. Weekes has argued that:

This reliance on "broad common sense" is perplexing. Expert evidence is tendered for the purpose of proving or disproving matters which are beyond the ordinary knowledge, sense and experience of the tribunal of fact. Thus when causation is beyond the powers of even the experts, the inferential approach amounts to deliberately uninformed guesswork.⁵⁷

So it is one thing to say that 'the law's task is to determine the nature and form of evidence the plaintiff should be permitted to adduce to prove the required historical link',⁵⁸ but it is not the law's task to find a link where the evidence is clear that the link cannot be established. As

⁵⁶ *Fairchild* (n13) [100] (Lord Hutton).

⁵⁷ Robert Weekes, 'Not Seeing the Wood for the Trees – Risk Analysis as an Alternative to Factual Causation in *Fairchild*' (2003) 12 Nott LJ 18, 25.

⁵⁸ Jane Stapleton, 'Publication Review – Causation and risk in the law of torts: scientific evidence and medicinal product liability' (2000) 116 LQR 506, 506.

Hoffmann said in *Fairchild*, ‘however robust or pragmatic the tribunal may be, it cannot draw inferences of fact in the teeth of the undisputed medical evidence’.⁵⁹

2.2 The ‘doubling of the risk’ approach to proof on the balance of probabilities

One final aspect of the balance of probabilities standard of proof must be addressed and that is the suggestion that the balance of probabilities can be satisfied by proof that the negligence more than doubled the background risk so is the ‘most probable’ cause. This issue was raised in the mesothelioma context in *Sienkiewicz v Greif*.⁶⁰ The claim was brought by the daughter of a mesothelioma victim who had been exposed to asbestos due to the negligence of the defendant employer. Unlike the claimant in *Fairchild* she had not been exposed to asbestos by any other employers, but she had been exposed to asbestos in the atmosphere of the area where she lived. This ‘environmental exposure’ was not attributable to negligence; it was an ‘innocent’ source of asbestos. The risk created by the environmental exposure was assessed at 24 cases per million, and the risk created by the negligent exposure was assessed at 4.39 cases per million, so the environmental exposure was by far the greater source of asbestos exposure. The defendant advanced two arguments. First it was argued that the *Fairchild* test should not apply in this case because, unlike *Fairchild*, there was only one negligent source of asbestos dust so it was suggested that the case did not involve the same degree of uncertainty as *Fairchild*. The standard but-for test should be applied, and, it was argued, causation could be proved on the balance of probabilities only by proof that the negligent exposure had more than doubled the risk created by the environmental exposure. Alternatively, the defendant argued that if the *Fairchild* test was applicable in the circumstances, the increase in risk attributable to the defendant’s negligence should only be regarded as ‘material’ if it at least doubled the background risk. The Supreme Court held that the *Fairchild* test did apply because the evidentiary gap concerning the aetiology of mesothelioma still exists and makes proof of causation impossible where there is more than

⁵⁹ *Fairchild* (n13) [70] (Lord Hoffmann).

⁶⁰ *Sienkiewicz* (n18).

one source of asbestos regardless of whether those sources are innocent or due to negligence. The evidentiary gap was not avoided by the potential to measure the extent of the risk attributable to each source.

Lord Phillips concluded in *Sienkiewicz* that statistical and epidemiological evidence was not sufficient to establish causation on the balance of probabilities, explaining that it was considered both inadequate and unreliable.⁶¹ The reliability of the epidemiological evidence is a factor that affects the degree of belief in the conclusions that can be drawn from it. This issue was addressed in chapter three, and although the statistical evidence in *Gregg* did not appear to be reliable it is important to remember that where reliable statistical and epidemiological evidence is available it can be a very valuable tool in proving causation. The issue that is specific to the evidentiary gap is the inadequacy of the epidemiological evidence. Lord Phillips suggested that ‘so long as medical science is unable to demonstrate, as a matter of fact, the aetiology of mesothelioma, data relating incidence to exposure is not a satisfactory basis for making findings of causation’.⁶²

It will be argued that Lord Phillips is right to say that epidemiological evidence cannot provide adequate proof of causation where there is an evidentiary gap, but that his criticisms go too far. It is important to observe that epidemiologists do not take evidence of a doubling of the risk as proof of a causal connection so rejecting the doubling of the risk approach does not entail rejection of epidemiological evidence. Indeed, once more is known about the aetiology of the disease, epidemiological evidence may be very valuable in proving causation.

2.2.1 Doubling of the risk

While there is still an evidentiary gap in relation to mesothelioma it is arguable that epidemiological evidence will continue to be inadequate proof of a causal link between the individual claimant and defendant where there are multiple sources of asbestos exposure. As

⁶¹ *ibid* [97]-[98] (Lord Phillips).

⁶² *ibid* [97] (Lord Phillips).

already explained, the evidentiary gap pertains to the aetiology of the disease and specifically to what constitutes a ‘sufficient set’. Because it is not possible to define a sufficient set of conditions to cause mesothelioma it is not possible to prove that the negligence was a necessary element of an actually occurring sufficient set of conditions. The balance of probabilities requires the court to form a belief in the fact of causation, and unless it is possible to state what fact is being proved, it is not possible to form a rational belief in that fact. As Steel and Ibbetson explain:

The balance of probabilities rule is still a probabilistic rule in so far as the fact-finder’s belief does not have to be a certain one. The main point, however, is that the probabilistic aspect of the rule is a reflection of the fact-finder’s degree of belief, not the content of that belief. This might be thought to be a purely philosophical point, but its consequences are clear: if statistical evidence only generates belief in a *probability*, not a fact, it cannot on its own satisfy the balance of probabilities rule.⁶³

This can be illustrated by analysing Beever’s statistical approach to *Fairchild*. He argued that the causal link between the claimant and a particular defendant/a particular source of asbestos could have been proven on the balance of probabilities in *Fairchild*, but his argument is reliant on statistical probabilities/relative frequency and, as seen above, this is not synonymous with proof on the balance of probabilities which denotes belief probability i.e. is qualitative. Beever illustrates his approach to the probability of causation in evidentiary gap cases through a variation on the single-hit hunters case of *Cook v Lewis*.⁶⁴ He introduces an element of evidentiary uncertainty by proposing an example where there are three equally likely possibilities: the hunter was hit by *A*, or was hit by *B* or was hit by both *A* and *B*. This is similar to the evidentiary gap since the causal mechanism may be a single bullet or it may be ‘cumulative’ if it involved both bullets. He says that the probability that *A* alone shot the hunter was 33.33 percent, the probability that *B* alone shot the hunter was 33.33 percent, and the probability that *A* and *B* both

⁶³ Steel and Ibbetson (n23) 464.

⁶⁴ [1952] 1 DLR 1

shot the hunter was 33.33 percent. This means that the overall chance that A shot the hunter is 66.67 percent (achieved by adding together the probability that it was A alone and the probability that it was A and B together). The same conclusion applies to B . Beever argues that ‘[t]he difficulty in *Fairchild* was caused, not by any inadequacy in the law’s traditional approach to causation, but by misunderstanding the medical evidence and failing to understand this feature of probability’.⁶⁵

Beever engages in a forceful defence of this mathematical aspect of the calculation of probabilities and his exposition of the mathematics is not challenged here. Unfortunately, as discussed above in relation to Barnes’ paper on the different types of probabilities, this purely mathematical probability is simply not what the balance of probabilities standard of proof involves. The law is concerned to establish the fact that A shot the hunter, not the mathematical probability that A shot the hunter. Unless the underlying causal process is known to involve either a single bullet *or* an accumulation of bullets, it is not possible to form a rational belief that A ’s bullet was *actually* causally relevant. It is essential to understand what constitutes a ‘sufficient set’ before being able to believe that A ’s bullet was necessary for an actually occurring sufficient set. This is part of the qualitative aspect of the balance of probabilities as a belief probability – the mathematical probabilities are clear but it is not possible to extrapolate to the individual case with any degree of credibility because of the evidentiary gap surrounding the aetiology.

Since Beever’s probability is not based on an understanding of the aetiology, it does not give rise to a high degree of belief in the soundness of his statistical probability. Miller has explained clearly that since the balance of probabilities relates to the degree of belief in a fact, where a court is faced with statistical evidence it must form a degree of belief in the credibility of that statistical evidence:

BOP [balance of probabilities] relates to the credibility or the degree of belief that the fact-finder ascribes to the claimant’s overall attempt to attribute her harm to the

⁶⁵ Beever (n38) 476.

defendant's negligence. If there is uncertainty in the scientific evidence, that attribution may have to be expressed in terms of risk, chance, odds, likelihood or other probabilistic terminology. But the fact-finder must still gauge the overall credibility of that account – thus, he may be 90% sure that the increased risk of the claimant's injury is less than 30%.⁶⁶

In the hunters scenario Beever provided, the mathematical probability that *A* was a cause was 66.67 percent, and the probability that *B* was a cause was also 66.67 percent. Despite this, if the claimant brought a case against either one of them, this mathematical probability would not be enough to create a belief that it is more probably than not true that he injured the claimant. This is because the claimant must first prove that it was either a single bullet or two bullets that hit him (what is a sufficient set – one bullet or two?), and then provide evidence that the defendant's bullet was necessary for the set that actually occurred.

2.2.2 Closing the evidentiary gap

The evidentiary gap may be overcome in the future as scientific knowledge develops so it is important to note that it will not be a permanent problem. Once the evidentiary gap no longer exists courts should revert to orthodox principles of damage and causation. McIvor criticises the decision in *Sienkiewicz*, observing that many of the speeches suggest that 'a claimant in a mesothelioma case would always need to be able produce evidence which identifies the precise point in time at which the genetic process triggering the malignant mutation was initiated'.⁶⁷ This, she says, reflects a standard of proof much closer to the criminal law requirement of proof 'beyond all reasonable doubt' than the civil law balance of probabilities standard. It is not necessary to prove the precise point in time when the mesothelioma was triggered, but it is necessary to prove that the particular causal process was more likely than not the process that

⁶⁶ Chris Miller, 'Liability for negligently increased risk: the repercussions of *Barker v Corus UK (Plc)*' (2009)8 LP&R 39, 51.

⁶⁷ Claire McIvor, 'Debunking some judicial myths about epidemiology and its relevance to personal injury litigation' (unpublished conference paper, Obligations VI: Challenging Orthodoxy conference, Canada 18th-20th July 2012) 12.

occurred. This is currently not possible; as Lord Phillips explained, while the single fibre theory has been discredited it is still not known whether the fibres have a synergistic interaction, nor at what stages in the development of the disease the fibres play a part. It is not possible to show that one process is more likely than the others. But if/when science enables us to say that causal potency does depend on the quantity/type of asbestos, or on intensity of exposure, or that earlier/later exposures are more likely to cause mesothelioma, then it may be possible to prove causation *on the balance of probabilities*.

McIvor explains that epidemiology is more far-reaching than Lord Phillips suggests it to be. In *Sienkiewicz* he stated that:

Epidemiology is the study of the occurrence and distribution of events (such as disease) over human populations. It seeks to determine whether statistical associations between these events and supposed determinants can be demonstrated. Whether those associations if proved demonstrate an underlying biological causal relationship is a further and different question from the question of statistical association on which the epidemiology is initially engaged.⁶⁸

But as McIvor argues, '[w]hile statistics are certainly an important tool used by epidemiologists, they are no more than a tool and epidemiologists are more than just statisticians'.⁶⁹ She explains that epidemiologists are trained not only in statistical techniques but also in research methodologies and medicine, so not only do they design and carry out studies, they then 'use various sophisticated techniques for interpreting the data precisely in order to determine whether any indicated statistical associations provide evidence of an underlying biological relationship of cause and effect'.⁷⁰ As she suggests, if epidemiologists were used as expert witnesses they would be able to draw meaningful and often useful conclusions from epidemiological evidence. Miller has also criticised the courts for being excessively dismissive of epidemiological evidence.

⁶⁸ *Sienkiewicz* (n18) [80] (Lord Phillips).

⁶⁹ McIvor (n67) 15.

⁷⁰ *ibid.*

Because the evidence is expressed in statistical or probabilistic terms, he says that courts have been unwilling to draw causal inferences but, he argues, the Bradford Hill criteria used by epidemiologists to assess the reliability of causal inferences ‘are unashamedly pragmatic and if judges were to reject epidemiological evidence which was deemed to have satisfied those criteria, then such opposition would be hard to defend’.⁷¹

This section has shown that where there is an evidentiary gap surrounding the aetiology of a disease which prevents the determination of a ‘sufficient set’ of conditions, and there are multiple sources of the harmful agent, it is not possible for a claimant to prove on the balance of probabilities that a particular source of the relevant harmful agent was a cause. Since the content of a sufficient set is unknown, even in theory, it is not possible to show that the defendant’s negligence was necessary for the sufficiency of any set of conditions that actually occurred. The claimant is required to prove the fact of causation, and the fact-finder must believe this to be more probably than not true in the sense that he must believe in the credibility of the evidence. The balance of probabilities standard of proof relates to *belief probability* which involves a qualitative assessment of the *fact probability*, of the *sampling error probability*, taking into account the design and conduct of the scientific study in question and of how far this can be extrapolated to the individual case of the claimant. The balance of probabilities has to be a belief probability because the fact-finder has to ask the question of the extent to which a study based on a given population can inform us about an individual with the claimant’s characteristics in the claimant’s circumstances.

This means that the fact-finder’s belief would be irrational if he chose to disregard clear scientific evidence as to causation and replace it with a ‘common sense’ view drawing on uninformed intuition. It also means that one cannot translate statistical probabilities directly on to the balance of probabilities standard of proof, so proof of a ‘doubling of the risk’ or a purely mathematical

⁷¹ Chris Miller, ‘Causation in personal injury: legal or epidemiological common sense?’ (2006) 26 LS 544, 566.

probability calculation such as Beever advocates, can only give rise to a probability of causation and the fact-finder must still ask what that actually tells him about the individual case. Without a scientific understanding of a sufficient set, one can only hold a low belief in the capacity of such statistical probabilities to provide meaningful information about an individual case. This means that the leap made by the majority in *McGhee* from finding that the defendant's negligence had 'materially increased the risk of harm' to concluding that his negligence had made a material contribution to the harm itself, cannot be accepted as drawing an inference about factual causation. As Nolan has argued, '[t]his conflation of material increase in risk and material contribution was a fairly blatant judicial sleight of hand'.⁷² Indeed, this interpretation was disapproved in *Fairchild* where it was held that the evidentiary gap prevented proof of causation, but that to assist claimants facing this insuperable problem of proof the defendant would be held liable on the basis that he had caused a material increase in the risk of harm. Under *Fairchild*, as developed in *Barker*, proof of a material increase in the risk of harm does not prove that the negligence was a cause of the harm. Instead the basis of liability has changed from liability for causing the claimant's loss, to liability for increasing the risk of the claimant's loss. Liability is no longer based on the fiction of a causal link, but explicitly based on the fact that the defendant increased the risk and this is enough to engage liability. The following section will therefore turn to this solution to the evidentiary gap and its place within a corrective justice-based account of negligence.

3. Risk of harm as the gist of the negligence action: *Barker v Corus (UK) plc*⁷³

In order for negligence liability to be consistent with corrective justice, there must be a causal link between the defendant's negligence and the claimant's loss. The previous section has shown that the evidentiary gap prevents proof of a causal link between the negligence and the physical

⁷² Donal Nolan, 'Causation and the Goals of Tort Law' in Andrew Robertson and Tang Hang Wu (eds.) *The Goals of Private Law* (Hart Publishing 2009) 168.

⁷³ *Barker v Corus (UK) plc* [2006] UKHL 20, [2006] 2 AC 572.

harm suffered by the claimant. The majority of the Supreme Court decision in *BAI v Durham* held that the *Fairchild* rule is a special rule whereby a defendant whose negligence materially increased the claimant's risk of mesothelioma is deemed to have caused the mesothelioma.⁷⁴ However, an alternative approach is to reconceptualise the actionable damage in these cases as exposure to the risk of harm rather than the physical harm itself since the claimant is able to prove that the defendant's negligence caused this 'loss'. This was the approach Lord Hoffmann purported to follow in *Barker v Corus*, and in his dissent in *BAI v Durham* Lord Phillips maintained that this was the correct interpretation of *Barker*. It will be argued here, however, that this approach is still not consistent with corrective justice. There are two strands to this argument. First, Lord Hoffmann insisted that the risk must have materialised into physical harm before the claimant would be able to recover for the exposure to the risk of harm, and this requirement means that it is still the physical harm that forms the gist of the action. Secondly, even if the physical harm requirement were to be removed, risk exposure lacks the independent moral significance to form damage in the negligence inquiry.

3.1 The decision in *Barker v Corus*

The claim in *Barker v Corus* raised the question of whether the *Fairchild* principle was applicable when the claimant had been exposed to asbestos not only due to the negligence of his former employers including the defendant, but also by himself during a period of self-employment. In *Fairchild* the House of Lords had said that any background exposure to asbestos was so small that it could be effectively discounted so all of the claimant's exposure was attributable to former employers.⁷⁵ This meant that, although the evidentiary gap prevented the claimant from proving that a particular defendant was a cause of his mesothelioma, it was accepted that his loss was certainly attributable to *somebody*'s negligence. The claimant in *Barker* was not an 'innocent

⁷⁴ *BAI v Durham* ("Trigger Litigation") (n4) (Lords Mance, Kerr, Clarke and Dyson).

⁷⁵ The question of whether it was correct to ignore the background risk, and the related 'innocent claimant' argument, will be addressed in section 4.1.

claimant' in this sense because the period of self-exposure meant that he could not say that his loss had definitely been caused by somebody else's negligence. Given the possibility that the claimant had caused his own loss the fairness of shifting the burden of the loss onto a negligent employer was less apparent. As an alternative argument, the defendant claimed that if the *Fairchild* principle was applicable in this case then liability ought to be several rather than joint and several.

The House of Lords held that the *Fairchild* principle did apply even though the claimant himself was responsible for some of the asbestos exposure. The court did, however, accept the defendant's second line of argument and held that where the *Fairchild* principle is applied liability ought to be several rather than joint and several. Lord Hoffman, with whom Lord Scott, Lord Walker and Baroness Hale agreed on the applicability of the *Fairchild* principle, based his finding on causal principles:

For this purpose it should be irrelevant whether the other exposure was tortious or non-tortious, by natural causes or human agency or by the claimant himself. These distinctions may be relevant to whether and to whom responsibility can also be attributed, but from the point of view of satisfying the requirement of a sufficient causal link between the defendant's conduct and the claimant's injury, they should not matter.⁷⁶

This echoes the argument made in Chapter Two that causation is a factual question of involvement so it is irrelevant to the causation inquiry whether a candidate cause is innocent or negligent. Instead this characterisation is relevant to the attribution of responsibility, of which causal involvement is just one element. In particular, the duty of care and principles of legal causation define the limits of moral responsibility.

Lord Rodger's dissent in *Barker* focused instead on balancing the interests of the claimant and defendant to achieve a just solution. He considered that the *Fairchild* principle had been motivated by the demands of fairness and the need to avoid injustice to the claimant. Assuming

⁷⁶ *Barker* (n73) [17] (Lord Hoffmann).

that liability remained joint and several, he was concerned that applying the *Fairchild* principle where the claimant was responsible for a period of asbestos exposure would tip the balance too far in his direction, resulting in injustice to the defendant.⁷⁷ However, given that the majority favoured several liability he accepted that ‘the balance of potential injustices favour applying the *Fairchild* exception’.⁷⁸

Section Four will look at broader policy arguments and considerations of fairness. The aspect of the decision that is considered here is the reasoning behind the decision that liability ought to be several. The House of Lords were in agreement that the extent of the defendant’s contribution to the total risk was an appropriate measure for the apportionment of liability that would take place. It remained a requirement that the claimant must have suffered physical harm in the form of mesothelioma, meaning that the risk must have materialised. The value of the total risk was therefore said to be the same as the value of the physical harm. The defendant would compensate the claimant for a portion of the value of the physical harm, with the extent of that portion measured by his contribution to the total risk. Lord Scott explained that this apportionment would be based on the duration of the exposure for which each defendant was responsible compared with the total duration of exposure to asbestos. It might also take into account the intensity of the exposure and, since different types of asbestos increase the risk more than others, the type of asbestos dust.⁷⁹

The majority were divided in the reasons they gave for this decision. One approach was to retain the physical harm, mesothelioma, as the actionable damage or ‘gist’ of the negligence action, but to apportion liability according to contribution to risk on the basis that this achieved a fair balance between the competing interests of the claimant and defendant. The second approach was to hold that where the *Fairchild* principle is applied the gist of the action is no longer the

⁷⁷ ibid [98]-[101] (Lord Rodger).

⁷⁸ ibid [101] (Lord Rodger).

⁷⁹ ibid [62] (Lord Scott).

physical harm but the exposure to risk itself. Baroness Hale favoured the first justification, that apportionment was a ‘fair’ solution:

The law of tort is not (generally) there to punish people for their behaviour. It is there to make them pay for the damage they have done. These *Fairchild* defendants may not have caused any harm at all. They are being made liable because it is thought fair that they should make at least some contribution to redressing the harm that may have flowed from their wrongdoing. It seems to me most fair that the contribution they should make is in proportion to the contribution they have made to the risk of that harm occurring.⁸⁰

The relevance of risk, in her approach, was no more than that it provided a ‘sensible basis’ on which liability could be apportioned.⁸¹

In contrast, risk took on much greater significance in Lord Hoffmann’s justification for applying several liability. Under the *Fairchild* principle a defendant is held liable on the basis that his negligence caused a material increase in the risk of harm. This meant, in his opinion, that the increase in risk, rather than the physical harm, ought to form the gist of the negligence action:

Consistency of approach would suggest that if the basis of liability is the wrongful creation of a risk or chance of causing the disease, the damage which the defendant should be regarded as having caused is the creation of such a risk or chance.

He continued to explain:

If that is the right way to characterize the damage, then it does not matter that the disease as such would be indivisible damage. Chances are infinitely divisible and different people can be separately responsible to a greater or lesser degree for the chances of an event happening.⁸²

This solution has an appearance of consistency with corrective justice since the defendant is ostensibly being held liable only for the loss that he has been proved to have caused i.e. exposure

⁸⁰ ibid [127] (Baroness Hale).

⁸¹ ibid [126] (Baroness Hale).

⁸² ibid [35] (Lord Hoffmann).

to the risk of mesothelioma. If the actionable damage is the risk of mesothelioma then the centrality of the causation requirement is maintained since the claimant must still prove on the balance of probabilities that the defendant's breach of duty caused this damage. On closer inspection, however, this approach is not consistent with corrective justice. There are two reasons for this. First, the continued insistence that the claimant must have developed mesothelioma undermines the claim that the gist has been changed to the risk of mesothelioma. In other words, the damage for which the claimant was compensated was still the mesothelioma itself, but liability was apportioned to reflect the degree of uncertainty surrounding proof of the causal link. Secondly, even if the courts were to abandon the requirement that the claimant must suffer the physical harm, exposure to risk cannot be considered to be damage within corrective justice because it adds nothing to the breach inquiry. Conduct is characterised as wrongful if it exposes the claimant to an unreasonable risk of harm. Corrective justice does not punish this wrongful behaviour, but requires the wrongdoer to repair the loss when his wrongdoing causes harm to another. If a defendant commits a wrong by exposing a claimant to an unreasonable risk of harm but this risk does not materialise then he is not liable because there is no loss for him to correct. If the law was to hold that exposure to risk does constitute harm then the requirement of damage would effectively be lost and the basis for liability would not be corrective justice but a retributive form of justice.

3.2 Analysis of the 'risk as gist' approach

As noted above, Lord Hoffmann explained that '[c]onsistency of approach would suggest that if the basis of liability is the wrongful creation of a risk or chance of causing the disease, the damage which the defendant should be regarded as having caused is the creation of such a risk of chance'.⁸³ Even Lord Rodger, who ultimately did not support the risk-based analysis, thought that the 'dominant aim' pursued by changing the gist of the action to the risk of harm was 'to

⁸³ *ibid* [35] (Lord Hoffmann).

secure internal consistency between the basis of liability and the nature of the damage'.⁸⁴

However, writing after the House of Lords' decision in *Fairchild*, Stapleton has said:

The majority of the House...firmly rejected the use of legal fictions. Refreshingly, they did not adopt any of a range of possible pretences. One would have been that the *McGhee/Fairchild* rule related to a special duty, such as a duty to avoid exposing the victim to an increased risk of mesothelioma which would have constituted a special type of wrong twinned with a special sort of 'damage'.⁸⁵

It will be argued that, as Stapleton suggests, the 'risk as damage' approach rests on a fiction and conceals the fact that the gist is actually still the physical harm and that the apportionment exercise reflects the remaining doubt over causation.

It will also be argued that Lord Hoffmann's approach introduces confusion to the interaction between the damage, causation, and quantification elements of the negligence inquiry. Stapleton has said that it 'cannot be over-emphasised that the formulation of the "damage" forming the gist of the action *defines* the causation question. Logically one can only deal with causation after one knows what the damage forming the gist of the action is'.⁸⁶ This is problematic in Lord Hoffmann's approach because he suggests that the gist of the action is 'risk' which is a divisible damage, but the claim is only actionable when the claimant has suffered the physical outcome harm, mesothelioma, which is indivisible. Where damage is divisible there should be 'apportionment' of liability as a matter of causal responsibility since each source of the harmful agent causes a portion of the total loss. Where damage is indivisible 'apportionment' (vis-à-vis the claimant; under joint and several liability the defendant can seek contribution from his co-defendants) takes on a different meaning. This exercise would take place at the liability stage as an exception to the usual rule of joint and several liability. The location of this 'apportionment'

⁸⁴ *ibid* [84] (Lord Rodger). His aim was to maintain 'a consistency of approach with the main body of law on personal injuries' (*ibid* [84]).

⁸⁵ Stapleton, 'Lords a'leaping' (n24) 290.

⁸⁶ Jane Stapleton, 'The Gist of Negligence: Part 2 The Relationship Between "Damage" and Causation' (1988) 104 LQR 389, 393.

exercise is therefore confused in Lord Hoffmann's approach which notionally treats a divisible harm as the gist while insisting on the occurrence of an indivisible harm and using this as a measure for the divisible harm.

3.2.1 The inconsistency of the physical harm requirement

Although Lord Hoffmann said that the gist of the action was the risk of mesothelioma, it was still a condition of liability that the claimant must have developed mesothelioma. This requirement of physical harm may be a practical way of limiting the number of cases where the *Barker* principles applies, but it is conceptually problematic because it means that the gist was not actually redefined as the risk of harm. Lord Hoffmann stated:

Although the *Fairchild* exception treats the risk of contracting mesothelioma as the damage, it applies only when the disease has actually been contracted. [Counsel for the defendant] was reluctant to characterise the claim as being for causing a risk of the disease because he did not want to suggest that someone could sue for being exposed to a risk which had not materialised. But in cases which fall within the *Fairchild* exception, that possibility is precluded by the terms of the exception. It applies only when the claimant has contracted the disease against which he should have been protected.⁸⁷

This reasoning is inadequate to support his assertion that the gist could be redefined as the risk of harm whilst still requiring the claimant to have developed mesothelioma. If the gist of the action is exposure to risk but the risk must have materialised before the action can be brought then there is an internal inconsistency between the content of the rule and the scope of application of the rule. Far from justifying the risk as damage approach, this inconsistency undermines the rule. Indeed, Beever says that 'the refusal to compensate the defendant's former employees unless they suffer mesothelioma reveals that the actionable damage is the mesothelioma and the consequences thereof, and not the risk of mesothelioma...The idea that

⁸⁷ *Barker* (n73) [48] (Lord Hoffmann). Affirmed in *Rothwell v Chemical and Insulating Co Ltd* [2007] UKHL 39, [2008] 1 AC 281.

these cases involve liability for risk creation is an illusion'.⁸⁸ The requirement that the claimant must have developed mesothelioma undermines the ‘risk as damage’ approach so severely that it cannot simply be explained away by Lord Hoffmann as a requirement for the application of the *Fairchild* principle. If he was willing to adopt the ‘risk as damage’ approach because he prioritised ‘consistency of approach’ then the physical harm requirement should have been abandoned in order to actually achieve consistency of approach. As noted by Scherpe, ‘[n]othing is gained by replacing one fiction with another but much can be lost’.⁸⁹

3.2.2 Physical harm and risk: a shift in perspectives

There is a danger that this argument seems to focus solely on the *appearance* of inconsistency that arises when the court says that the gist is the risk of mesothelioma whilst still insisting that the claimant must have developed mesothelioma. Examining the nature of ‘risk’, however, will show that this problem is not just apparent, but goes to the essence of the concept of risk. The following section will look more closely at what ‘risk’ is, to show that risk is a forward-looking concept and that, properly understood, the extent of the risk created by a particular defendant is measured at the time it is created. There is an inherent mismatch between risk as a forward-looking concept and the backward-looking role in which risk is being cast in Lord Hoffmann’s approach. Here the extent of the defendant’s contribution to risk is only measurable once the risk has materialised and all the other sources of the same kind of risk have been identified. This means that the extent of the risk that the defendant created will vary in size depending on how many other sources of asbestos the claimant was exposed to, whereas risk is actually a forward-looking concept and measureable independently of other sources of risk. This mistaken use of risk seems attributable to the subtle shift in the phraseology of the *Fairchild* exception which

⁸⁸ Beever (n38) 487. See also Nolan (n72) 178; Lara Khoury ‘Causation and risk in the highest courts of Canada, England and France’ (2008) 124 LQR 103, 126.

⁸⁹ Jens Scherpe, ‘A New Gist?’ (2006) 65 CLJ 487, 488.

originated as ‘material increase in risk’ in *McGhee* and has morphed into ‘material contribution to risk’ in *Barker*.

3.2.2.1 What is ‘risk’?

Perry has said that ‘[i]n ordinary language conduct is typically said to be risky when it gives rise to a chance of a bad outcome of some kind. The concept thus involves two main elements: first, a notion of chance or probability, and second, a notion of harm’.⁹⁰ It involves a state of uncertainty as to the future. The measure of a particular risk is a product of the extent of the possible harm that may be produced, and the probability of this harm materialising. Notably, probability is a factor in calculating risk but it is not synonymous with risk. In *Risks and Legal Theory*, Steele set out a range of different meanings assigned to the word ‘risk’. She explained that for some theorists we can only truly speak of ‘risk’ when the probability of the relevant outcome can be calculated. If the probability cannot be calculated then the situation ought to be described as involving uncertainty, indeterminacy, or ignorance. Others use the word ‘risk’ to describe any situation involving a hazard regardless of whether the probability of harm can be calculated. She says that the latter approach ‘reflects an increasing colloquial and theoretical understanding of risks as *threats*, rather than as statistical probabilities’.⁹¹ Notably the notion of risk relates to the future in all of these forms. ‘Risk’ is also sometimes used as shorthand for a particular decision-making method. In this sense ‘risk’ does not refer to a danger that necessitates a decision-making method in order to be managed, but refers to a way of approaching a problem or danger.⁹² Steele also notes that ‘risk’ and ‘probability’, ‘though related, are distinct terms’.⁹³ Risk will sometimes depend on a formal assessment of probability, and always depends on the likelihood of an outcome. Probability can be used for analysis of past events, e.g. what is the probability that *X*

⁹⁰ Stephen R Perry, ‘Risk, Harm, and Responsibility’ in David G Owen (ed.) *Philosophical Foundations of Tort Law* (Oxford University Press 1995) 322.

⁹¹ Jenny Steele, *Risks and Legal Theory* (Hart Publishing 2004) 7.

⁹² *ibid* 7.

⁹³ *ibid* 19.

caused Y ?), so probability does not always entail a risk assessment. In contrast, risk generally concerns the future although it can be used to debate the past,⁹⁴ for example to assess what would in the past have been a rational way to behave.⁹⁵ Once again we see that probability is one element of risk, but it is not synonymous with risk. This means that it is crucial that the two terms, ‘risk’ and ‘probability’, are not used interchangeably.

So while probability may be used to analyse past events, risk is a forward-looking concept. It describes a situation of uncertainty as to whether a particular outcome will occur. When a risk materialises it causes an outcome so it is no longer a ‘risk’ but a cause. Of course there may be uncertainty as to what caused a particular event i.e. uncertainty as to which risk(s) materialised, and as Steele noted above, probability can aid the analysis of causation. It would be incorrect to say that if a risk has materialised it has contributed to the risk of the outcome – this is just circular. In other words a ‘risk’ is only a ‘risk’ while the outcome is still prospective. Once the outcome occurs then the relevant question is whether the risk materialised and made a causal contribution. Weekes explains:

By definition the concept of risk is a limitless one. The meaning of a factual cause is binary: a given activity either causes or does not cause the damage in question. The limit of a factual cause is where a given activity does not result, in whole or in part, in the certain damage. Risk, however, is defined by degree. A risk which is proven to have resulted in damage is of course a risk that has been realized, or more properly, “a cause”.⁹⁶

Beever explains the same idea through the analogy of a raffle that is held with 100 tickets each costing £2 and a single prize of £1,000 (where the ticket has no value other than the fact it provides a chance to win the £1,000 prize). Before the raffle is drawn, he explains, ‘your ticket is

⁹⁴ ibid 20.

⁹⁵ ibid 9.

⁹⁶ Weekes (n57) 27.

worth £10, that being onehundredth of £1,000. But imagine now that the raffle is held and you do not win. How much is your ticket worth now? It is worthless?⁹⁷

3.2.2.2 Can risk constitute damage?

Lord Hoffmann held in *Barker* that mesothelioma is no longer the actionable damage for which the claimant recovers. Instead the actionable damage for which the claimant recovers is the risk of mesothelioma to which the defendant exposed him. However, the claimant cannot claim in respect of this exposure to the risk of mesothelioma until he actually develops mesothelioma, and on Lord Hoffmann's reckoning it is then possible to establish how much the defendant contributed to the total risk to which the claimant was exposed. It will be argued that what this approach actually achieves is to make the defendant liable for the mesothelioma itself, but to discount the extent of his liability to reflect the uncertainty over whether the risk he created was the risk that actually materialised. The method used to calculate the appropriate discount is the *probability* that it was the defendant's risk rather than another source of risk that materialised. It cannot be the case that exposure to risk is the damage that is being compensated here because risk is forwards-looking and only exists before the outcome occurs. After the outcome has occurred, the risk that the defendant created either is or is not a cause of that outcome. Since risk is forwards-looking the extent of the risk is measurable at the time of exposure to the risk and its value is fixed at that time. In contrast, in Lord Hoffmann's approach, the measure of the defendant's liability will vary with each subsequent exposure of the claimant to the same type of risk. This variability shows that it simply cannot be 'risk' that is actually being compensated in Lord Hoffmann's approach. Instead he is awarding partial compensation for the mesothelioma to reflect the *probability* that the defendant caused it. But we must remember that 'probability' is *not* synonymous with 'risk' – probability is one element of risk calculation, but whereas

⁹⁷ Beever (n38) 486

probability can look both forwards and backwards, risk only looks forwards. This is the argument that will be elaborated in more detail here.

3.2.2.2.1 Calculating risk and probability

As explained, risk is a forward-looking concept and is a product of the extent of possible harm and the probability of that harm occurring. When an employer negligently exposes an employee to asbestos dust he exposes the employee to the risk of a number of asbestos-related diseases such as lung cancer, asbestosis and mesothelioma. The risk can be measured at that time based on the extent of the loss the employee would suffer if the risk were to materialise and the probability of it materialising. If the employee chose to take out a health insurance policy it would take into account the increased risk of the employee suffering these illnesses in the future – it is the increase in the probability of the disease that we tend to focus on because the extent of the loss does not vary. This focus on the probability aspect of risk may explain why Lord Hoffmann uses the term ‘risk’ when he is actually referring to ‘probability’.

The relevant risk in these cases is the risk of mesothelioma. To provide a concrete example, in *Sienkiewicz* the risk of mesothelioma created by the environmental exposure to asbestos was 24 cases per million. The risk of mesothelioma created by the occupational exposure to asbestos was 4.39 cases per million.⁹⁸ The occupational exposure was therefore said to have increased the risk of the claimant developing mesothelioma by 18 percent.⁹⁹ ‘Risk’ here is a predictive tool that informs us how likely it is that the employee will develop mesothelioma. Based on the duration and extent of the exposure to asbestos dust, and the type of dust, it tells us how much more likely it is that the employee will develop mesothelioma than if she had not been exposed to each source of asbestos dust. If, hypothetically, the claimant had then been exposed to the same kind of risk by a later employer this would not alter the fact that the defendant had increased the risk

⁹⁸ For the purposes of this example it is assumed that these figures are accurate although McIvor notes that the judge calculated them himself from limited evidence and without the help of epidemiologist expert witnesses (n67).

⁹⁹ $(4.39 / 24) * 100 = 18.29$

of mesothelioma by 18 percent. The risk that was created by the defendant does not increase or diminish when the employee is exposed to the same kind of risk by subsequent employers.

This can be explained through an equivalent variation on Beever's raffle analogy. In that example 100 tickets were sold for a raffle with a single prize of £1,000. If *A* has one ticket, his ticket is worth £10 since that is one hundredth of £1,000. In terms of 'risk', the value of the 'risk' is £10 because it is a product of the value of the possible outcome, £1,000, and the probability of that outcome, 1/100. If *B* gives his ticket to *A*, he has added another 1/100 to the probability that *A* will win £1,000. If *C* has two tickets and gives these to *A*, he has added another 2/100 to the probability that *A* will win £1,000. *A*'s initial ticket is still worth £10 because it has a 1/100 chance of winning, the ticket he got from *B* is still worth £10 because it has a 1/100 chance of winning.

Lord Hoffmann said that 'the basis of liability is the wrongful creation of a risk or chance of causing the disease'.¹⁰⁰ If this was the case then the value of the risk created by a particular defendant would be fixed at the time of the exposure to that risk and would not vary. But this is not how Lord Hoffmann's approach works, so liability cannot actually be being imposed in respect of risk. Starting with the raffle analogy, the equivalent to developing mesothelioma would be *A* winning the raffle prize of £1,000. The 99 losing tickets are worthless; the 1 winning ticket is worth £1,000. In the first scenario where *A* only has one ticket, this is the ticket that has been drawn and Lord Hoffmann would say that *A* contributed the whole of the 'risk' of winning. In the second scenario where *A* had another ticket given to him by *B*, Lord Hoffmann would say that *A* contributed 50 percent of the 'risk' of winning, and *B* contributed the remaining 50 percent. And in the final scenario where *A* was given a further two tickets by *C*, Lord Hoffmann would say that *A* contributed 25 percent of the 'risk' of winning, *B* contributed 25 percent of the 'risk' of winning, and *C* contributed 50 percent of the 'risk' of winning. So the extent of the 'risk' created by *A* varies, as does the extent of the 'risk' created by *B*. But we know that risk is a

¹⁰⁰ Barker (n73) [35] (Lord Hoffmann).

forwards-looking concept and that its measure does not depend on other subsequent exposures to the same kind of risk, so Lord Hoffmann cannot actually have made ‘risk’ the actionable damage. Once again the problem appears to be that he treats ‘risk’ as synonymous with ‘probability’ perhaps because of the role that probability plays in allowing us to calculate risk. As noted above, we also have a tendency to focus only on the probability element of risk when comparing risks of the same outcome because the value of the possible outcome is a constant and the risk-creator increases the probability of this outcome occurring. Since the outcome has occurred, *A* has won £1,000, one of his tickets is worth £1,000 and the remaining three are worth nothing. There is a $1/4$ probability that it is his own ticket, a $1/4$ probability that it is the ticket he received from *B* and a $2/4$ probability that it is a ticket he received from *C*. If we made him divide up the prize according to these probabilities, keeping £250 for himself, giving £250 to *B* and £500 to *C* we would be allocating the prize according to the purely statistical *probability* that each person was responsible for the winning ticket. It may be that in cases of evidentiary gap this is the best the court can achieve, to apportion liability according to the statistical probability that the defendant caused the disease. But since causation of the physical outcome cannot be proven on the balance of probabilities, the decision to impose proportionate liability for possible causation cannot be based on corrective justice. If we want to recognise exposure to the risk of mesothelioma as actionable damage in itself, then we need to understand that this damage occurs at the time of the exposure and has value that is independent of other exposures to the same kind of risk.

3.2.2.2.2 ‘Increase in risk’ or ‘contribution to risk’?

This conflation of risk and probability is hidden in the subtle, yet significant, shift in the wording of the *Fairchild* principle. The *Fairchild* principle is expressed both as a test of ‘material increase in the risk’ of harm, and of ‘material contribution to the risk’ of harm. These phrases seem similar

but they actually mean very different things. The Court of Appeal in *Williams v University of Birmingham* asked whether:

...the law requires a judge to do a comparison between the two (or more) sources to see if the wrongful exposure at the hands of the defendant is, by comparison with the other exposures, too insignificant to be taken into account. Or, is it enough for a judge to find as a fact that the wrongful exposure at the hands of the defendant materially increased the risk that the victim would contract mesothelioma?¹⁰¹

Asking whether the defendant's negligence materially increased the risk of harm reflects the forwards-looking concept of risk. An increase in the risk is measurable at the time of the exposure and if it is a 'material' increase then this will not change because the measure of the risk is unaffected by later exposures. In contrast, asking whether the defendant's negligence made a 'material contribution to the risk' does require a comparison with the other sources of risk. The contribution that the defendant made to the total 'risk' will be greater or smaller depending on how many other employers exposed the claimant to asbestos, over what period and with what intensity. As discussed above, this variability in the measure of the so-called 'risk' shows that actually we have shifted to measuring probability of causation. This change in the terminology seems to have occurred in academic discussion following *Fairbeld*.¹⁰² It was firmly entrenched by the time of *Sienkiewicz* where Lord Phillips said 'the only circumstances in which a Court will be able to conclude that wrongful exposure of a mesothelioma victim to asbestos did not materially increase the victim's risk of contracting the disease will be where that exposure was insignificant compared to the exposure from other sources'.¹⁰³ Whether conscious or unconscious, it is possible that it reflects a desire to draw a parallel with the *Wardlaw* test of 'material contribution to harm' that is a standard test for causation. The discussion above shows that courts must be

¹⁰¹ [2011] EWCA Civ 1242, [2012] PIQR P4 [69] (Aikens LJ).

¹⁰² See for example Stapleton 'Lords a'leaping' (n15).

¹⁰³ *Sienkiewicz* (n18) [11] (Lord Phillips). Although the Court of Appeal in *Williams* concluded that it does *not* require a comparative exercise (n101) [72] (Aikens LJ).

careful when making subtle changes to the wording of a legal test because this can open the way for the test to take on a very different meaning.

3.2.3 Confusion in the ‘apportionment’ issue

Beever has highlighted another fallacy in Lord Hoffmann’s approach, explaining that if the ‘harm’ is divisible then responsibility for it is already ‘apportioned’ as a matter of causation without the need subsequently to apply several liability.¹⁰⁴ Lord Hoffmann’s approach is therefore problematic because he suggests that the gist of the action is ‘risk’ which is a divisible damage, but the claim is only actionable when the claimant has suffered the physical outcome harm, mesothelioma, which is indivisible. This means that he collapses the distinction between defining the damage that the defendant caused and determining the extent of the defendant’s liability for that damage.

As argued in Chapter One, it is important to maintain a clear distinction between the various elements of the negligence inquiry. Stapleton has explained that the question of the extent of responsibility is distinct from the question of causal responsibility, and different terminology applies to each issue. Where damage is divisible and the defendant caused only a portion of the total damage, he is only causally responsible for that portion, so may only be held liable for that portion. Stapleton says:

Though it is a rough usage it is convenient to call this causal ‘apportionment’ to distinguish it from the notion of ‘proportionate’ liability which refers to the, rarely adopted, rule that a tortfeasor is only liable to the claimant for a proportion of an indivisible injury as well as distinguishing it from the notion of ‘contribution’ between tortfeasors.¹⁰⁵

¹⁰⁴ Beever (n38) 488. See for example *Holtby v Brigham & Cowan (Hull) Ltd* [2000] 3 All ER 421 (CA), discussed at Ch 2 (n75) and (n94).

¹⁰⁵ Stapleton, ‘Lords a’leaping’ (n15) 283.

In contrast, in cases of indivisible damage, liability cannot be ‘apportioned’ on the basis of causal responsibility. Instead liability must be divided on the basis of non-causal considerations of comparative responsibility, and this is properly called ‘several’ liability. This means that if the damage in *Barker* really was the risk of mesothelioma, which is a divisible harm, the defendant has only caused a distinct portion of the overall risk exposure. Since he is causally responsible for only a portion of the total risk, legal liability can only arise in respect of this portion of the total risk, so there is no need to then apply several liability. As Beever explains, ‘[t]he defendant created the whole of the relevant risk. It does not make sense to speak of attributing any share of that risk to him’.¹⁰⁶ The fact that a kind of ‘apportionment’ was achieved through several liability shows that it is still the mesothelioma itself that forms the gist of the action.¹⁰⁷ This is also evident in the fact that the ‘risk’ is valued as a portion of the outcome harm rather than having intrinsic value of its own.¹⁰⁸

So far it has been shown that requiring the claimant to have developed mesothelioma before allowing him to bring a claim means that the gist of the negligence action has not been redefined as the risk of mesothelioma. The focus will now turn to the question of whether risk could form the gist of a negligence action if the physical harm requirement was to be abandoned.

3.3 Can risk constitute damage in corrective justice?

This section focuses on risk solely in the forward-looking sense that the term properly implies. The question is whether exposing somebody to the risk of mesothelioma or of other harm is capable of constituting damage within a corrective justice-based system of negligence law. In this forward-looking sense it is still important to draw a distinction similar to that in the previous chapter on loss of a chance. In that chapter it was seen that causal processes can be deterministic

¹⁰⁶ Beever (n38) 488.

¹⁰⁷ Note that the Supreme Court in *Sienkiewicz* (n18), and the majority in *BAI v Durham* (n4), confirmed that the gist of the action is the mesothelioma.

¹⁰⁸ The question of the value of a risk will be considered below.

or indeterministic.¹⁰⁹ It is only in indeterministic processes that we can appropriately say that there is a ‘chance’ of an outcome occurring and that the probability of this outcome is ‘objective’. In deterministic processes, as most physical processes are believed to be, there is not a ‘chance’ as such since the occurrence of a particular outcome can theoretically be predicted with certainty given limitless knowledge. Instead the probability used to describe the likelihood of the outcome occurring is an epistemological probability, it reflects our belief based on the limited knowledge that is available. Thus a patient who is unwell when he visits his doctor is either destined to recover, or destined to succumb, he does not have an objective chance of recovery. Generally speaking then we deal with epistemological and not objective probabilities.

Recalling that probability is one element of risk, along with the possible outcome, risk can be objective or subjective even though it is based on epistemological probability. The term ‘risk’ denotes uncertainty as to the likelihood of a future outcome, so it incorporates epistemological probability. The harmful agent/risk agent has the potential to cause an outcome i.e. it has the potential to form part of a deterministic causal process, but the limits of our knowledge prevent us from saying whether it will or will not cause harm to a particular individual. But there is a limit to this because the relevant outcome has an objective nature. With an indivisible harm such as mesothelioma, one only creates an objective risk of mesothelioma if the potential victim has not already contracted the disease. Of course, an employer will not usually know whether the employee has already contracted mesothelioma, so from his perspective exposing the employee to asbestos exposes the employee to a risk of mesothelioma, so in this subjective/epistemic sense the conduct creates a risk of harm. So even though risk relies on epistemological *probability*, the objective nature of harm means that we can talk about objective *risk* and subjective/epistemic *risk*. The following sections will consider whether risk, in either sense, can constitute damage in the negligence inquiry.

¹⁰⁹ See Ch 3 section 4.

3.3.1 Objective risk: the evidentiary gap prevents proof that the defendant created an objective risk

A criticism of the ‘risk as damage’ approach that has emerged in academic writing is that the evidentiary gap also prevents the claimant from proving that the defendant’s negligence increased the risk that he would develop mesothelioma.¹¹⁰ Scherpe has explained:

The aetiology of the disease is such that once contracted further exposure does not matter and certainly cannot increase the risk of contracting it. If the disease was contracted during the first employment, all following exposure did and could not increase the risk and hence there could not be a contribution to the risk. But whether that was the case, we do not know – which is exactly the point: we do not know whether all exposure actually increased the risk. Assuming that it did is therefore also resorting to a fiction.¹¹¹

In other words, it is only before mesothelioma has been contracted that exposure to asbestos actually creates an objective risk that the individual will develop the disease. Since the evidentiary gap prevents us from saying when the disease was contracted it is also impossible to say that any exposures other than the first exposure actually created a risk of the disease for this individual.

Nolan has addressed a potential criticism of this argument:

It could be argued that this is to adopt the wrong perspective. From the point of view of the defendant at the time of the breach of duty, a risk *was* imposed on the claimant...and indeed this must always be the case where the defendant has been negligent *towards the claimant*, as he has to be in order to be liable. But even if we assume that this was the perspective the House of Lords had in mind when they said that in cases falling within the *Fairchild* principle, the defendant was deemed to have caused the injury when he could be shown to have materially increased the risk of its happening, this switch from objective risk to subjective risk cannot rescue the risk as damage idea. This is because

¹¹⁰ See Nolan (n72) 178; Beever (n38) 486; Scherpe (n89) 488.

¹¹¹ Scherpe (n89) 488.

even if we accept the premise that an unrealised risk is a form of harm, this claim can surely be sustainable only if the defendant's action did *in fact* create a risk for the claimant.¹¹²

Since the evidentiary gap prevents us from being able to say whether the defendant exposed the claimant to an objective risk of mesothelioma, risk in the objective sense cannot help overcome the problem of the evidentiary gap.

3.3.2 Subjective/epistemic risk: lacks the moral significance to be damage

The only remaining sense in which risk may therefore constitute damage, if it is to assist the courts in overcoming the evidentiary gap, is in the subjective sense. As explained above, a lack of knowledge will often prevent an employer from knowing whether his employee has already developed mesothelioma, so exposing the employee to asbestos exposes him to a risk of mesothelioma in a subjective, or epistemic sense. This is what enables the conduct to be characterised as wrongful, or negligent.¹¹³ Since the standard of care in negligence is an objective standard, for the avoidance of confusion the label 'epistemic' will be used instead of 'subjective'.

As noted above, Nolan argued that while the *Fairchild* principle may be based on the fact that from the defendant's perspective he exposed the claimant to the risk of mesothelioma this does not rescue risk as a form of damage. It will be argued that Nolan's assertion is correct and that epistemic risk cannot constitute damage in negligence. One barrier is the difficulty of explaining why this sort of risk matters and should be deserving of compensation. The other, far greater, barrier is that it adds nothing to the negligence inquiry. Corrective justice requires a wrongdoer to repair the loss that he caused the victim to suffer through his wrongdoing. If the concept of 'loss' is the same as the concept of wrongdoing, i.e. the creation of an unreasonable risk, then loss is effectively subsumed into the concept of wrongdoing and liability is imposed for the sole fact of wrongdoing towards the claimant. An entire aspect of the equation has been removed,

¹¹² Nolan (n72) 179.

¹¹³ Clearly asbestos exposure is also wrongful because it creates a risk of other asbestos-related illnesses too.

and the focus is solely on the defendant's wrongdoing so the justification for liability can no longer be corrective justice but must be a defendant-focused form of justice such as retributive or distributive justice.

3.3.2.1 The difficulty of explaining why epistemic risk deserves compensation

In terms of the negligence doctrines, Stapleton has argued that the question of what constitutes 'gist damage' or actionable damage is 'rarely addressed squarely by courts', and that the word 'damage' is 'bandied about in a number of different contexts, usually without clear definition yet equally without apparent awareness of the importance of precision in its use'.¹¹⁴ This highlights one of the key difficulties with accepting risk as damage in cases involving the evidentiary gap – should exposure to the risk of harm be considered damage for the purposes of corrective justice-based liability? This question was not addressed head-on by the House of Lords. Amirthalingam has therefore argued that '*Barker* is an unsatisfactory decision in that it does not explain why "increased risk" should qualify as the gist of negligence'.¹¹⁵ Furthermore, it does not account for why the risk of mesothelioma, and possibly of other diseases whose aetiology involves the same kind of evidentiary gap, is considered to be damage but the risk of other physical harm is not regarded as damage. It will be argued that any limit on what kinds of risk qualify as damage will necessarily be arbitrary because of the more fundamental problem that the risk of harm simply cannot constitute damage. Each of these objections must be addressed in greater detail.

A significant barrier to accepting that the risk of harm constitutes damage is that it is difficult to say why risk has value and what that value is. Voyiakis explains that 'we cannot decide whether persons exposed to risk of physical harm should be entitled to claim compensation unless we have some rough idea of what they would demand to be compensated *for*'.¹¹⁶ As discussed in the

¹¹⁴ Jane Stapleton, 'The Gist of Negligence: Part 1 Minimal Actionable Damage' (1988) 104 LQR 213, 213.

¹¹⁵ Kumaralingam Amirthalingam, 'The Changing Face of the Gist of Negligence' in J Neyers, E Chamberlain & S Pitel (eds.) *Emerging Issues in Tort Law* (Hart Publishing 2007) 474.

¹¹⁶ Emmanuel Voyiakis, 'The Great Illusion: Tort Law and Exposure to Danger of Physical Harm' (2009) 72 MLR 909, 915.

previous chapter, the risk itself does not have a value as a proportion of the value of the physical harm to which it relates:

[W]hen I make it 40 per cent more likely that you will suffer lung cancer, I am not causing you 40 per cent of the harm that lung cancer brings about. The quantity of risk and the quantity of physical harm do not seem to be related in any such straightforward way. We therefore have reason to object to the latter being taken as a measure of the former.¹¹⁷

Instead, Voyiakis focuses on the concrete effects that exposure to the risk of harm can create, notably preventative medical care and the increase in the cost of health insurance or life insurance. Steele has distinguished the ‘utility value’ of risk from its ‘mathematical value’ and explained that the utility value of any particular risk depends on the individual involved. She says that ‘if the possible losses are *not* purely financial, it is much harder to attach the risk ‘value’ in any calculable form at all’.¹¹⁸ So even the question of whether a risk has a financial impact will depend on the individual and whether they are inclined to have health/life insurance anyway.

Similarly, risk cannot be valued according to the anxiety it causes the claimant. The House of Lords held in *Rothwell* that the risk of future asbestos-related illness is not actionable damage, nor is anxiety related to this risk, nor is the combination of the two.¹¹⁹ Green supports this conclusion, noting that human beings necessarily operate within a notion of risk meaning that:

the degree to which any individual might or might not actually be at risk of an adverse outcome need bear no resemblance whatsoever to his *perception* of that risk. Therefore, any anxiety founded upon such a necessarily epistemic conception of that risk is a step further removed from the defendant’s breach of duty than is anxiety consequent upon actual harm resulting from the tort.¹²⁰

¹¹⁷ ibid 917.

¹¹⁸ Steele (n91) 25.

¹¹⁹ *Rothwell* (n87).

¹²⁰ Sarah Green, ‘Risk Exposure and Negligence’ (2006) 122 LQR 386, 388.

It is also difficult to reconcile risk as damage with the decision in *Cartledge v E Jopling & Sons Ltd*.¹²¹ The claimant there had suffered damage to his lungs caused by inhalation of silica dust and it was held that the fibres caused actionable damage before producing symptoms that were noticeable in everyday life. If physical damage is actionable before it is discoverable by the claimant then risk exposure ought to also be actionable regardless of whether the claimant is aware of the risk. Yet risk exposure only seems to acquire significance if the claimant is aware of it. If the claimant is unaware of the risk then he is unaffected unless it materialises in physical harm.

The argument that epistemic risk should be regarded as damage could be seen to be analogous with the argument that was made in the previous chapter that the loss of an epistemic chance should be regarded as damage in the limited context of medical negligence involving the misdiagnosis or mistreatment of existing illness. That argument, however, was based on the unique characteristic of the doctor's duty of care which is focused on the illness, and the fact that both the patient and doctor regard the patient as having a 'chance' of recovery. There is no such focus in this context. Amirthalingam has noted, therefore, that limiting recovery for risk to mesothelioma, or to risks arising in the course of employment, would be arbitrary – there is nothing marking these contexts out as ones in which the claimant and defendant regard the risk of harm as damage in advance of their interaction. Amirthalingam suggests that 'unacceptable risk' might be considered damage. He explains that this 'provides a normative framework for courts to determine what sort of risks should be classed as unacceptable, and incrementally develop the law'.¹²² This solution is too vague to be applied in practice, and it still does not explain *why* a particular risk is unacceptable.

¹²¹ [1963] AC 758 (HL).

¹²² Amirthalingam (n115) 475.

3.3.2.2 Epistemic risk: the conflation of breach and damage

The more fundamental obstacle to the recognition of epistemic risk as damage in negligence is that the damage requirement would be subsumed into the breach inquiry so liability would be based entirely on wrongdoing rather than on the interaction of the two parties, so it would no longer be a system of corrective justice.

In *Fairchild*, Lord Rodger suggested that '[a]t best, it was only good luck if any particular defendant's negligence did not trigger the victim's mesothelioma'.¹²³ This is true, but should not impact on corrective justice-based liability since this addresses wrongful *loss*. As discussed in Chapter One,¹²⁴ the causation requirement has been criticised as introducing an element of 'moral luck' into negligence liability,¹²⁵ and if the focus of liability was the wrongdoing of the defendant alone then there may be some force to this argument. However since the focus of corrective justice is not on either of the parties alone, but is on the inequality that occurs when the wrongdoing of one person causes a loss to another person, it is only concerned with equality in *transactions*. When two people interact there is only a transaction when something passes between them i.e. one causes a loss to the other. As Steele has explained, 'if the defendant is lucky and does not cause damage, then potential victims are lucky too and do not suffer any'.¹²⁶ The claimants in these cases have not been 'lucky' overall because we know that they have developed mesothelioma but this does not mean that it was caused by the particular defendant. If the defendant was lucky and did not trigger the claimant's mesothelioma but simply created a risk of mesothelioma by exposing him to asbestos dust then there has been no corrective justice transaction between the defendant and the claimant.

The relationship between the *Fairchild* test and breach of duty was considered by the Court of Appeal in *Williams v University of Birmingham*.¹²⁷ The main issue to be decided in that case was

¹²³ *Fairchild* (n13) [155] (Lord Rodger).

¹²⁴ See Ch 1 text to (n88).

¹²⁵ David Howarth, 'Many duties of care – or a duty of care? Notes from the underground' (2006) 26 OJLS 449, 461.

¹²⁶ Steele (n91) 116.

¹²⁷ *Williams* (n101).

whether the defendant had breached his duty of care towards the claimant. The judge at first instance said that the legal standard of care was ‘to take all reasonable measures to ensure that [the claimant] was not exposed to a material increase in the risk of mesothelioma’.¹²⁸

In the Court of Appeal, however, it was held that this was inaccurate:

A reference to exposure “*to a material increase in the risk of mesothelioma*” brings the test for causation in mesothelioma cases into the prior questions of the nature of the duty and what constitutes a breach of it. There is nothing in *Fairchild* or [*Sienkiewicz v Greif*] to suggest that the House of Lords or the Supreme Court has altered the “breach of duty” test in mesothelioma cases so that a claimant only has to demonstrate that the defendant failed to take reasonable steps to ensure that the claimant or victim was not exposed to a “*material increase in the risk of mesothelioma*”.¹²⁹

Instead the duty of care is ‘to take reasonable care...to ensure that [the claimant] was not exposed to a foreseeable risk of asbestos related injury’.¹³⁰ If the defendant did not materially increase the risk of harm then it was clear that he had not breached his duty of care. But if his conduct did materially increase the risk of harm then the court must still ask whether a reasonable person in the defendant’s place would have taken further steps to reduce the risk of harm. The test of breach of duty is therefore more demanding than the test of material increase in the risk of harm that is applied at the causation stage, so it is clear that epistemic risk is already subsumed into the breach inquiry. Subsuming the damage requirement into the breach inquiry would mean that the only remaining elements of the negligence inquiry would be the duty of care and the breach of duty. The negligence action would focus almost exclusively on the defendant, and would no longer be a system of corrective justice.

¹²⁸ (Belcher J) at first instance (unreported) cited in *Williams* (CA) (n101) [39]

¹²⁹ *Williams* (CA) (n101) [40] (Aikens LJ).

¹³⁰ *ibid* [40].

4. Theoretical approaches to the ‘evidentiary gap’

In this chapter it has been argued that the ‘evidentiary gap’ presents an insurmountable obstacle to proof of causation. This cannot coherently be overcome by taking a more relaxed evidential approach and drawing a common sense inference of causation. Nor can it be avoided by moving the problem to a different aspect of negligence by reconceptualising the gist of the action. This final section therefore turns to the question of how to justify imposing liability, whether it is several or joint and several, when causation cannot be proven. The dominant theme in judicial reasoning in *Fairchild* and in *Barker* is that the solutions were justified by policy concerns and the demands of fairness and justice. It will be argued here that ‘fairness’ and ‘justice’ are not arguments of corrective justice. The first part seeks to clear away a number of possible misconceptions about this. In the second part it will be demonstrated that they reflect an instrumentalist approach to liability, although the exact content of these instrumentalist goals is not clearly articulated in the decisions. As a result of adopting an instrumentalist approach, the *Fairchild* principle lacks a coherent theoretical foundation since, as discussed in Chapter One, instrumentalist theories cannot be properly implemented through the institution of negligence law. As Nolan has explained, ‘[t]he trouble is that the logic of the instrumentalist approach has no obvious stopping point...And because there are no logical stopping points – or rather because the logical stopping points are incompatible with the basic idea of tort law - the courts have had to resort to illogical ones instead’.¹³¹

4.1 Dispelling possible misconceptions

First it is important to address a number of possible misconceptions that ‘fairness’ may have some claim on corrective justice.

¹³¹ Nolan (n72) 189.

4.1.1 McGhee: the defendant definitely caused the loss

In *Fairchild*, Lord Bingham noted that the Court of Appeal had sought to distinguish that case, where multiple former employers may have caused the disease, from *McGhee* where there was only one employer involved. He said that in *McGhee*, ‘there was a risk that the defendant might be held liable for acts for which he should not be held legally liable but no risk that he would be held liable for damage which (whether legally liable or not) he had not caused’.¹³² This, however, is not a corrective justice-based reason for imposing liability. In Chapter One it was seen that a similar argument arose in relation to *Chester v Afshar*.¹³³ There the defendant doctor had negligently failed to warn the patient about a particular risk involved in an operation, and the risk did in fact materialise even though the operation itself was carried out without negligence. The patient may still have undergone the operation if she had been properly warned, so she would still have been exposed to the risk that materialised, which meant that the failure to warn her of the risk was not a cause of her loss. The defendant doctor was held liable despite the absence of the causal link. In order to achieve this outcome, however, the House of Lords had to rely on distributive arguments such as the importance of vindicating the patient’s right to give informed consent, and the need to deter doctors from breaching their duty to warn patients of risks.¹³⁴ In corrective justice, a person is only required to repair loss that is caused *by their wrongdoing*. If loss was caused by non-wrongful, i.e. non-negligent, conduct then it is not a wrongful loss. It is therefore irrelevant that the defendant employer in *McGhee* was responsible for all of the claimant’s exposure to brick dust. Some of this exposure was negligent but some was innocent and unless it can be shown that the negligent exposure specifically was a cause of the claimant’s dermatitis then the loss is not wrongful.

¹³² *Fairchild* (n13) [33] (Lord Bingham). See also Joe Thomson, ‘*Barker v Corus: Fairchild chickens come home to roost*’ (2006) Edin LR 421, 424.

¹³³ [2004] UKHL 41; [2005] 1 AC 134. See Ch 1 text to (n85).

¹³⁴ See Ch 1 section 1.5.1

4.1.2 *Fairchild*: the ‘innocent claimant’ argument

In *Fairchild*, the House of Lords considered that any background risk of mesothelioma was so small it could be effectively discounted, so that it was possible to say that the claimant’s harm had definitely been caused by the negligence of *a* former employer it was just impossible to say *which* former employer(s) in particular.¹³⁵ Lord Bingham said:

[T]here is a strong policy argument in favour of compensating those who have suffered grave harm, at the expense of their employers who owed them a duty to protect them against that very harm and failed to do so, when the harm can only have been caused by breach of that duty.¹³⁶

Assuming for now that any non-negligent exposure could properly be disregarded, this would mean that the claimant could prove that he had suffered a ‘wrongful loss’ because he could prove that his loss had been caused by somebody else’s negligence, so the claimant’s side of the corrective justice equation would be fulfilled. This is similar to the problem that arose in the American case of *Sindell v Abbott Laboratories*.¹³⁷ The claimants were young women who had developed a cancer which was caused by their mothers having taken the drug Diethylstilbestrol (DES) during pregnancy. They were unable to prove which particular manufacturer had produced the drug that their mother had taken. Goldberg has said that ‘factual causation was not disputed in the DES cases, but what became an issue was the identification of the particular manufacturer of the product and a causal link between it and the damage’.¹³⁸ But as Beever explains:

We cannot say that the claimant is deserving of recovery and leave it at that. It is irrelevant to corrective justice that the claimant deserves recovery from *someone*...In

¹³⁵ *Fairchild* (n13) [2] (Lord Bingham).

¹³⁶ *Fairchild* (n13) [33] (Lord Bingham). See also [41] (Lord Nicholls), [153] (Lord Rodger).

¹³⁷ 26 Cal. 3d 588, 607 P 2d 924 (1980).

¹³⁸ Richard Goldberg, *Causation and Risk in the Law of Torts: Scientific Evidence and Medicinal Product Liability* (Hart Publishing 1999) 11.

corrective justice, it is necessary to say that the claimant is deserving of recovery *from some particular person*.¹³⁹

This is because corrective justice is not concerned with the claimant and defendant individually, but with both of them in their interaction, so it is not enough to say that the claimant has suffered a wrongful loss. He must be able to say that his wrongful loss was caused by the particular defendant's wrongdoing before that particular defendant can have any corrective justice-based *interpersonal responsibility* to repair the loss. It is important that the responsibility is individual as highlighted by Nolan. He argues that corrective justice cannot accommodate the focus on the claimant's wrongful loss because this 'refers not to each defendant as an individual but to the defendants as a collective entity'.¹⁴⁰ He explains:

[W]hile it seems reasonable to say that, as between the claimant and the *defendants*, the equities favour the claimant, unless there is some good reason why we should 'collectivise' the defendants in this way, we still do not have a justification for the imposition of liability as between the claimant and each individual defendant.¹⁴¹

Moreover, 'from the individual defendant's point of view, it seems irrelevant that the other possible cause of the claimant's injury is the negligence of another person, as opposed to non-negligent conduct or a natural event'.¹⁴² Even if the claimant has suffered a wrongful loss, in the sense that his loss was definitely caused by somebody else's negligence, corrective justice still only requires that loss to be repaired by the individual whose negligence caused it. There may be a strong reason to feel that the claimant is deserving of compensation, but if a particular defendant cannot be proved to have caused the loss then there is no corresponding reason to engage his personal responsibility.

The above discussion is premised on the assumption in *Fairchild* that all the sources of asbestos had been created by the negligence of former employers. This is not always the case. In *Barker*

¹³⁹ Beever, *Rediscovering* (n38) 457.

¹⁴⁰ Nolan (n72) 174.

¹⁴¹ *ibid* 174.

¹⁴² *ibid* 175.

the claimant had negligently exposed himself to asbestos for a period of time. More importantly, it seems that the court was wrong to ignore the background risk through environmental exposure. The majority of the claimant's exposure to asbestos in *Sienkiewicz* was not due to negligence but was present in the atmosphere of the area where she lived. Mesothelioma can be caused by very fleeting exposures to asbestos.¹⁴³ Nolan suggests that imposing liability on a defendant in these circumstances reflects an underlying assumption 'that somehow fault matters more than causation'.¹⁴⁴ But, as he explains:

[W]ithin a corrective justice framework this assumption is simply false, for, as Allan Beever points out, 'from the perspective of corrective justice, the defendant is *entirely innocent* (with respect to the claimant) unless the claimant establishes *all* [the elements of the cause of action] against the defendant'.¹⁴⁵

Imposing liability, whether joint and several or merely several, on a defendant whose negligence has not been proven to be a cause of the claimant's loss is simply not consistent with corrective justice-based interpersonal responsibility. Indeed, although the attempt to reformulate the gist of the action as the risk of harm in *Barker* was ultimately flawed, it did highlight the House of Lords' awareness that the causation requirement is an indispensable element of the negligence inquiry. Commenting on a similar approach in American tort law, Ripstein and Zipursky explain:

Some recent cases sidestep causation issues by redefining the category of injury that will permit recovery...All of these efforts tend to underscore, rather than to rebut, the centrality of causation in extant American tort law, as they indicate the need to shift to other categories, rather than abandoning causation itself.¹⁴⁶

While it may be argued that there is a clear need to compensate the claimants in these cases, and that it is important to deter defendants from acting negligently, it is not the role of negligence

¹⁴³ Stapleton, 'Lords a'leaping' (n15) 276-9.

¹⁴⁴ Nolan (n72) 175.

¹⁴⁵ ibid 175, citing Beever, *Rediscovering* (n38) 446.

¹⁴⁶ Arthur Ripstein and Benjamin C Zipursky, 'Corrective Justice in an Age of Mass Torts' in Gerald J Postema (ed.) *Philosophy and the Law of Torts* (Cambridge University Press 2001) 219.

liability to pursue these goals. The coherence of negligence law is dependent on its pursuit of corrective justice-based personal responsibility, so it cannot be adapted to plug the gaps in order to achieve these consequentialist goals. As seen in Chapter One, if the law is ‘stretched’ to achieve compensation and/or deterrence then it will suffer in terms of coherence and tend to be measured in terms of how well it achieves these consequentialist aims leading critics, such as Atiyah, to suggest that negligence law should be abandoned in favour of a fairer compensation system.¹⁴⁷ It is important to recall that while negligence law gives effect to corrective justice this does not preclude distributive goals such as compensation and deterrence from being pursued through other, more appropriate, institutions. As Ripstein and Zipursky argue:

[A]ll of the considerations put forward as reasons for attenuating the causation requirement in tort law are more plausibly viewed as reasons for supplementing the legal system with public compensatory and deterrent systems, for tort law without causation will be both ineffective and arbitrary to the core.¹⁴⁸

In the absence of state schemes it may seem unjust to leave the claimant without compensation in negligence, and to allow the negligent defendant to escape liability, but this is unjust from a distributive perspective and from the perspective of corrective justice liability must be denied. As explained in Chapter One, it is one thing to acknowledge that negligence liability has compensatory and deterrent effects, but it is another thing to say that it has compensatory and deterrent goals and to measure its success against those goals. As a system of compensation or deterrence it has been shown that negligence will necessarily be a failure because the goals are mutually truncating and incompatible with the correlative structure of the negligence action.¹⁴⁹ If we deny liability where causation is not proven then negligence law might be criticised for failing to achieve optimal compensation and deterrence but it is not appropriate to measure its success or failure on those terms. It would be like measuring a car in terms of how good it is as an

¹⁴⁷ See Ch 1 text to (n57) and (n58), P S Atiyah, *The Damages Lottery* (Hart Publishing 1997) 32-95.

¹⁴⁸ Ripstein and Zipursky (n146) 245.

¹⁴⁹ See Ch 1 text to (n54) and (n55).

aeroplane and criticising it because it is no good at flying – we are criticising it for failing to be something that it is not. Instead we are interested in how well negligence law succeeds at achieving corrective justice and, measured in those terms, negligence law is a success if it denies liability where causation cannot be proven.

4.2 Where next?

This final section turns to the question of how the courts should resolve future cases that involve an evidentiary gap. Since the passing of the Compensation Act 2006, these cases must be divided into two categories: those cases involving mesothelioma, and those involving other diseases.

4.2.1 Mesothelioma

The House of Lords decision to impose several liability in *Barker* was reversed, and joint and several liability was reinstated in cases involving mesothelioma by s.3 Compensation Act 2006. The provisions of s.3 Compensation Act 2006 apply where the defendant is liable in tort in connection with damage caused to the victim by the disease ‘whether by reason of having materially increased a risk or for any other reason’.¹⁵⁰ As Lord Phillips noted in *Sienkiewicz*, as and when more is understood about mesothelioma and the uncertainty surrounding the aetiology is removed the *Fairchild* test may no longer apply since there would be no need for an exceptional approach to causation if the evidentiary gap no longer existed.¹⁵¹ It is therefore essential that the courts assess future developments in the scientific evidence relating to the aetiology of mesothelioma. This chapter has shown the importance of using the NESS test to understand precisely which aspect of causation is affected by the evidentiary gap. It has also examined the relationship between scientific and legal standards of proof, and highlighted the need for judges to engage fully with scientific evidence so that they understand not only the *limits* to what it can prove but also the full extent of what it *is* capable of proving. If science progresses to a stage

¹⁵⁰ s.3(1) Compensation Act 2006.

¹⁵¹ *Sienkiewicz* (n18) [70] (Lord Phillips).

where it is able to describe a sufficient set of conditions to cause mesothelioma then the evidentiary gap will no longer exist. Any remaining difficulties of proof will relate to the necessity of the defendant's negligence for the sufficiency of an actually occurring set of conditions. This may be easier in some cases than others, but it will at least be notionally possible. Given that the *Fairchild* test is inconsistent with corrective justice, the courts should cease to apply it as soon as the evidentiary gap is closed.

4.2.2 Other diseases involving an evidentiary gap

The Compensation Act 2006 only applies to mesothelioma which means that if the *Fairchild* test is applied to other diseases involving an evidentiary gap the *Barker* principle of several liability is applicable. Although Lord Brown in *Sienkiewicz* suggested that the *Fairchild* exception should be limited to mesothelioma actions,¹⁵² it has not been convincingly restricted in this way so it may potentially be applied more widely.¹⁵³ This final section considers whether the *Fairchild* exception can and should apply to claims involving any other diseases where there is an evidentiary gap. It will be argued that the 'single harmful agent' rule adopted in *Wilsher* and approved in *Fairchild* provides a rational limit on when the *Fairchild* exception can apply. However, turning to the question of whether it should be applied outside the context of mesothelioma, it will be argued that since the *Fairchild* exception lacks a coherent theoretical justification it ought not to be applied to other diseases. Although the application of several liability following *Barker* is a 'fairer' solution than joint and several liability, this does not rescue the *Fairchild* exception from its lack of a theoretical basis.

¹⁵² ibid [187]

¹⁵³ Chris Miller, 'Causation in personal injury after (and before) *Sienkiewicz*' (2012) 32 LS 396, 410.

4.2.2.1 The single harmful agent rule: *Wilsher v Essex Area Health Authority*¹⁵⁴

As discussed in chapter two, the claimant in *Wilsher* was a baby born prematurely who later suffered blindness caused by retrolentalfibroplasia ('RLF') and his case was brought in the interval between *McGhee* and *Fairchild*. To recap the facts of this case, the defendant's negligence consisted of administering excess oxygen which carries a risk of RLF and the baby also suffered four other conditions due to being born prematurely, each of which also carried a risk of RLF. He was unable to prove on the balance of probabilities that the excess oxygen was a cause of his RLF using the but-for test. He therefore argued instead that the *McGhee* principle should be applied, and the defendant should be held liable on the basis that his negligence materially increased the risk of RLF. The majority of the Court of Appeal found that the *McGhee* principle was applicable because the aetiology of the claimant's RLF was unknown, and applying the *McGhee* test they found that the defendant's negligence had materially increased the risk of RLF. Browne-Wilkinson V-C dissented, and his dissenting opinion was approved in the House of Lords. He distinguished the facts in *Wilsher* because there was only one harmful agent in *McGhee* i.e. brick dust, whereas in *Wilsher* there were five possible harmful agents i.e. the excess oxygen created by the defendant and the four other naturally occurring conditions.

The defendants failed to take reasonable precautions to prevent one of the possible causative agents (e.g. excess oxygen) from causing RLF. But no one can tell in this case whether excess oxygen did or did not cause or contribute to the RLF suffered by the plaintiff. The plaintiff's RLF may have been caused by some completely different agent or agents... To the extent that certain members of the House of Lords [in *McGhee*] decided the question on inferences from evidence or presumptions, I do not consider that the present case falls within their reasoning. A failure to take preventative measures

¹⁵⁴ [1988] AC 1074 (HL).

against one out of five possible causes is no evidence as to which of those five caused the injury.¹⁵⁵

Since it could not be proved that excess oxygen was a cause of the claimant's RLF it was not possible to say that the defendant's negligence had led to a material increase in an operative risk:

[U]nless and until you can say that the plaintiff's RLF was caused by oxygen, it is impossible to say that the injury falls "squarely within the risk".¹⁵⁶

The House of Lords reversed the decision of the Court of Appeal in *Wilsher*, preferring Browne-Wilkinson's 'single harmful agent' rule.¹⁵⁷ Although Lord Hoffmann dissented on this point in *Fairchild*, arguing that the rule does not draw a 'principled distinction',¹⁵⁸ he later accepted a more nuanced form of the single harmful agent rule in *Barker* explaining:

In my opinion it is an essential condition for the operation of the exception that the impossibility of proving that the defendant caused the damage arises out of the existence of another potential causative agent which operated in the same way. It may have been different in some causally irrelevant respect...but the mechanism by which it caused the damage...must have been the same. So, for example, I do not think that the exception applies when the claimant suffers lung cancer which may have been caused by exposure to asbestos or some other carcinogenic matter but may also have been caused by smoking and it cannot be proved which is more likely to have been the causative agent.¹⁵⁹

The single harmful agent rule has been widely criticised. Nolan, for example, says that 'it is difficult to see any logic in this limitation',¹⁶⁰ and Beever says that it 'seems to be nothing more

¹⁵⁵ *Wilsher v Essex AHA* [1987] 1 QB 730 (CA), 779 (Browne-Wilkinson VC).

¹⁵⁶ *ibid* 780 (Browne-Wilkinson VC).

¹⁵⁷ It will be recalled that Lord Bridge held that *McGhee* had not laid down a new legal principle but that the court had taken a 'robust and pragmatic approach' to the facts in order to draw a legitimate inference of causation. The House of Lords in *Fairchild* subsequently overruled this aspect of Lord Bridge's reasoning, holding that *McGhee* had in fact laid down a new legal principle and that the inferential approach was inappropriate, but it affirmed the outcome in *Wilsher*, so the 'single harmful agent' rule still applies.

¹⁵⁸ *Fairchild* (n13) [72].

¹⁵⁹ *Barker* (n73) [24] (Lord Hoffmann).

¹⁶⁰ Nolan (n72) 173.

than an arbitrary restriction on the application of *Fairchild*.¹⁶¹ It will be argued that from a purely policy-driven perspective the rule is indeed arbitrary, but from a factual causation perspective it is justified because it limits the level of scientific uncertainty under which a defendant may still be held liable.

As this chapter has shown, the decision in *Fairchild* is inconsistent with the principles of corrective justice, and instead the House of Lords explicitly resorted to distributive justice-based arguments to justify their decision. Distributive arguments were also raised in *Fairchild* to rationalise the single harmful agent rule that had been established in *Wilsher*. Unlike *McGhee* and *Fairchild* which involved an employer's negligence, Lord Hoffmann said that in the medical context of *Wilsher* 'it cannot possibly be said that the duty to take reasonable care in treating patients would be virtually drained of content' unless the rules on causation are relaxed.¹⁶² Furthermore, 'the political and economic arguments involved in the massive increase in the liability of the National Health Service...are far more complicated than the reasons...for imposing liability upon an employer who has failed to take simple precautions'.¹⁶³ Viewed in these terms the single harmful agent rule is indeed arbitrary. If liability is imposed because of the need to enforce the defendant's duty of care,¹⁶⁴ or because of the need to compensate the claimant,¹⁶⁵ then it is irrelevant whether there was a single harmful agent involved.

The policy arguments about liability are, however, not the appropriate starting point for our reasoning. The *Fairchild* exception addresses a problem of proof of factual causation so it is essential to begin by identifying the exact nature of the causal problem before considering the appropriate solution. In terms of the NESS test for causation, it has been explained above that where there is an evidentiary gap it is impossible to describe a 'sufficient set' in general terms.

¹⁶¹ Beever, *Rediscovering* (n38) 475.

¹⁶² *Fairchild* (n13) [69] (Lord Hoffmann).

¹⁶³ *ibid* [69].

¹⁶⁴ *ibid* [33] (Lord Bingham).

¹⁶⁵ *ibid* [36] (Lord Nicholls).

But it is important here to specify precisely why this was impossible and also to restate the terms of the NESS test:

[A] particular condition was a cause of (condition contributing to) a specific consequence if and only if it was a necessary element of a set of antecedent *actual* conditions that was sufficient for the occurrence of the consequence.¹⁶⁶

In both *McGhee* and *Fairchild* it was known that the sufficient set that *actually* occurred contained the type of substance to which the defendant had exposed the claimant. The sufficient set that actually occurred in *McGhee* contained abrasions from brick dust, and in *Fairchild* it contained inhalation of asbestos fibres. What was unknown was *how* the particular substance forms a sufficient set – does a sufficient set contain just one or many abrasions/asbestos fibres? In *Wilsher*, there was an extra degree of uncertainty because it was impossible to say that the sufficient set that *actually* occurred even contained excess oxygen. A sufficient set *can* contain excess oxygen. A sufficient set might also contain excess oxygen in combination with one or more of the other conditions.¹⁶⁷

If the baby in *Wilsher* had not suffered the other conditions then it might be said that an evidentiary gap remains in relation to oxygen since Mustill LJ explained in the Court of Appeal that ‘[w]hat is no longer so clear as it was, is that the concentration rather than the duration of the exposure is crucial’.¹⁶⁸ It is not possible to say how oxygen causes RLF since there is uncertainty what ‘excess’ means. It may relate to the concentration of oxygen, or to the duration, or both. The court could thus apply the *Fairchild* test and say that a doctor who negligently exposed a baby to very high levels of oxygen for a short time/a slight excess for a long time, had increased the risk of RLF. But there is an extra level of uncertainty in the actual facts of *Wilsher* because it is not even possible to say that oxygen was part of the sufficient set *that actually occurred*. Given that it was very exceptional for the court to depart from corrective justice and allow the

¹⁶⁶ Wright, ‘Causation in Tort Law’ (n26) 1790 (emphasis added).

¹⁶⁷ *Wilsher* (CA) (n155) 764-5 (Mustill LJ).

¹⁶⁸ *ibid* 766.

claimant to leap the evidentiary gap in *Fairchild*, it would be even more exceptional and require even stronger justification to allow a claimant to leap an even wider evidentiary gap such as existed in *Wilsher*.

4.2.2.2 Limiting the *Fairchild* exception to mesothelioma

The decision in *Fairchild* was quite openly motivated by considerations of ‘fairness’ and ‘justice’ and other ‘policy’ considerations. Similarly in *Barker* the legal rules were adapted to achieve what was perceived as a fair solution, with Lord Walker going so far as to say ‘I prefer to start with the more fundamental issue of apportionment, since it must have a bearing on how far and how fast the boundaries of the new principle are to be extended’.¹⁶⁹ This openness has received some praise, with Morgan commenting that the ‘open acknowledgment [in *Fairchild*] that policy had to be considered was welcome’,¹⁷⁰ and Khoury suggesting that in *Barker*:

The House of Lords has arguably been more transparent and honest than the French Cour de cassation by recognising that proportional liability is, in such cases, imposed *without proof of causation* between the negligence and the *actual loss*, due, essentially, not to logical conceptual reasons, but to the existence of overriding policy objectives.¹⁷¹

But it is not sufficient merely to acknowledge openly that a decision departs from traditional principles because of policy concerns. In order to be justifiable, those policy concerns must be articulated and defended. Morgan thus criticised the *Fairchild* decision because ‘[t]he court must explain precisely which policies are at stake, and such analysis is missing in *Fairchild*.¹⁷² Broad reference to ‘fairness’ and ‘justice’ is also insufficient as Morgan explains:

¹⁶⁹ *Barker* (n73) [107] (Lord Walker). See also [40]-[43] (Lord Hoffmann), [101] (Lord Rodger), [109] Lord Walker, [127] (Baroness Hale).

¹⁷⁰ Jonathan Morgan ‘Lost Causes in the House of Lords: *Fairchild v Glenhaven Funeral Services*’ (2003) MLR 277, 279.

¹⁷¹ Lara Khoury, ‘Causation and Risk in the Highest Courts of Canada, England and France (2008) 124 LQR 103, 130.

¹⁷² Morgan (n170) 279.

While these are noble sentiments, they should not be dignified with the label of policy. Such a ‘policy’ says nothing more than that ‘injured claimants should recover’, and therefore it is much too wide to be of any use in setting the boundaries of what, after all, is said to be an exceptional approach...A convincing justification must explain why recovery is allowed in certain cases, *and in those cases only*.¹⁷³

The argument that it is unjust to leave claimants uncompensated is also criticised by Nolan who says ‘this position was not really fully articulated, no justification being given for the assertion that irredeemable evidential uncertainty should operate to the detriment of the negligent defendants rather than the blameless employee’.¹⁷⁴ It is also difficult to reconcile with the fact that under *Barker* there is a significant chance that the claimant will not be fully compensated if defendants can no longer be traced or have become insolvent. Kramer argues that ‘while it is clearly unfortunate that the victims are not fully compensated, arguably this is a result of tortfeasors becoming insolvent and not of the apportionment rule itself’.¹⁷⁵ But if the *Fairchild* exception was devised to achieve a fair outcome for claimants in the face of a practical problem of proof, surely this entire area of law is shaped by the desire to achieve a particular outcome in practice, so it does not make sense to ignore the practical impact of insolvency of defendants when designing the legal rule. As the decision in *BAI v Durham* attests, the decision in *Barker* has potential to impact on the interpretation of the terms of insurance contracts relating to the date that the injury was ‘sustained’ or ‘contracted’ because this may depend on whether the relevant damage is the risk or the physical harm.¹⁷⁶

It should be remembered that s.3 Compensation Act 2006 only applies to mesothelioma, so the *Barker* principle of apportionment would apply to other evidentiary gap disease cases. The decision to apply several liability in *Barker* seems fairer than joint and several liability, but what is fairness being measured against? Green has argued that ‘although there may well exist strong

¹⁷³ ibid.

¹⁷⁴ Nolan (n72) 172.

¹⁷⁵ Adam Kramer, ‘Smoothing the Rough Justice of the Fairchild Principle’ (2006) 122 LQR 547, 550.

¹⁷⁶ *BAI v Durham* (n4).

arguments in favour of maintaining the rule against apportionment where causation can be established on an orthodox basis, it is perfectly reasonable to argue for a modification of that rule where the causal criteria have themselves been altered'.¹⁷⁷ This shows the significance of the *Barker* decision – the abandonment of the causation requirement in *Fairchild* meant that liability was no longer based on corrective justice which meant that other negligence doctrines could also be altered to achieve a particular vision of liability. As such, Steel explains '[t]he real function of 'risk' in *Barker* is as a practical tool for devising apportionment of liability in the interests of fairness'.¹⁷⁸ Steel goes on to explain that '[s]uch a consideration of fairness is only intelligible against the background that there is a (moral) difficulty with the proposition that a defendant who has not caused the mesothelioma may be liable for it'.¹⁷⁹

One proposal for providing a rational limit to *Fairchild* is that it could be limited to cases of 'industrial risk', but it will be argued that this does not provide a sound theoretical basis for the rule. As noted above, Lord Hoffmann considered that it was appropriate to apply the *Fairchild* principle in a claim brought by an employee against a former employer but not one brought by a patient against their doctor. Indeed, Amirthalingam has argued that to avoid throwing causation 'into the abyss of intuition and policy...it would have been better if the case had been analysed squarely as one about liability for industrial (or unacceptable) risk'.¹⁸⁰ The difficulty with characterising the *Fairchild* exception as involving industrial risk is that this focuses on defendants to the detriment of the claimants. As discussed in Chapter One, deterrence is a distributive goal centred on defendants and it cannot be properly achieved through the institution of negligence law which involves claimants too. Given that liability in *Fairchild*, and under s.3 Compensation Act 2006, is joint and several, defendants are arguably being over-deterring. If deterrence is to be implemented through the negligence system which requires a claimant to bring an action then

¹⁷⁷ Sarah Green 'Winner takes all' (2004) 120 LQR 566, 568.

¹⁷⁸ Sandy Steel, 'Uncertainty Over Causal Uncertainty: *Karen Sienkiewicz (Administratrix of the Estate of Enid Costello Deceased) v Greif (UK) Ltd*' (2010) 73 MLR 646, 653. See also *Barker* (n73) [120]-[128] (Baroness Hale).

¹⁷⁹ Steel (n178) 653.

¹⁸⁰ Amirthalingam (n115) 470-1.

there ought to be a focus on the needs of the claimant too. But if *Fairchild* was a principle relating only to industrial risk then a claimant exposed to asbestos in school (or university) would be unable to bring a successful claim against the school or the workmen employed by the school.¹⁸¹ And although Amirthalingam's focus on the industrial context seems to emphasise the deterrent effect of liability, he also seems concerned by the vulnerability of the claimant in the industrial context because he favours joint and several liability (subject to reduction for contributory negligence) since this 'leaves the risk of insolvency of other potential defendants on the defendants'.¹⁸² As Chapter One demonstrated, the structure of negligence law prevents distributive goals from being achieved in a comprehensive way, and within a negligence claim the competing distributive arguments relate to the claimant and the defendant separately and are mutually truncating.

The difficulty with this is that 'fairness' is still not explained, beyond saying that some concession is made to both the claimant and to the defendant. Once fairness is prioritised over corrective justice in developing legal rules, the courts ought to consider what is fair to both claimants and defendants. This pluralism introduces incoherence because, as Chapter One showed, goals relating to the claimant and defendant individually will be mutually truncating.¹⁸³ This has been seen in other areas of negligence. In *Gwilliam v West Herts NHS Trust*,¹⁸⁴ the court imposed a duty of care on the defendant to take reasonable care to ensure that a contractor had public liability insurance. This duty did not arise because of considerations of interpersonal responsibility between the claimant and defendant but because of the court's desire to ensure that compensation could be paid. Morgan has criticised this decision because insurance is a distributive institution so while it may be desirable for it to operate alongside the negligence system it cannot drive the content of the legal principles. As he argues, '[t]he courts are

¹⁸¹ Such as occurred in *Knowsley Metropolitan Borough Council v Willmore* [2011] UKSC 10, [2011] 2 WLR 523, and *Williams v University of Birmingham* (n101).

¹⁸² Amirthalingam (n115) 474.

¹⁸³ See Ch 1, section 1.3.

¹⁸⁴ *Gwilliam v West Herts NHS Trust* [2002] EWCA Civ 1041; [2003] QB 443.

powerless to complete the change to a (rational) compensation scheme...Thus, they must ensure that tort continues to function in its own terms, namely the legal doctrinal argument which characterises private law'.¹⁸⁵ The same criticism applies to *Fairchild*.

Given that negligence law cannot fully achieve goals relating to either the claimant or the defendant, the decision about where the balance between the competing interests ought to lie is subjective. Morgan has highlighted this by contrasting the decisions of the Court of Appeal and the House of Lords in *Fairchild*:

The perceived unfairness of [leaving the claimant without compensation] had...not swayed the Court of Appeal...who set against the unpalatability of denying recovery to sympathetic claimants the injustice of requiring a defendant who may well have had no causal connexion with the relevant injury to compensate the claimant's entire loss...The House of Lords were content to assert that this injustice was 'plainly' outweighed by the injustice of the claimants recovering nothing, which reasoning is (equally plainly) contestable and certainly not much of an explanation.¹⁸⁶

With this in mind, the decision in *Barker* to apply several liability seems to be the fairest solution because it seeks to find a balance between the competing interests of the claimant and the defendant. But it must be made clear that this is a personal preference as to which distributive arguments ought to be prioritised over others in a context where such a choice simply has to be made and articulated clearly because of the underlying incompatibility of achieving distributive goals through a system of corrective justice. Given that proportionate liability cannot be justified by principles of corrective justice-based interpersonal responsibility it is preferable for the *Fairchild* exception to be restricted to claims in respect of mesothelioma.

¹⁸⁵ Jonathan Morgan, 'Tort, insurance and incoherence' (2004) 67 MLR 384, 400.

¹⁸⁶ Jonathan Morgan, 'Lost Causes' (n170) 281.

Conclusion

This chapter has shown that it is important to understand the different approaches in *Barker* in detail. On first impression, Lord Hoffmann's reconceptualisation of the gist of the action as the risk of harm appears to reconcile the *Fairchild* test with the principles of corrective justice; since the claimant can only prove that the defendant's negligence exposed him to the risk of harm he is only compensated for the risk exposure rather than the physical harm. The subtle change in the formulation of the *Fairchild* test also appears to bring it in line with the *Wardlaw* test of material contribution to harm, the only difference being that the 'harm' where *Fairchild/Barker* applies is 'the risk of harm'. In comparison, the reasoning adopted by Baroness Hale seems flimsy and unprincipled since it is based on vague notions of what she feels is a 'fair' balance between the competing interests of the claimant and defendant. This chapter has shown, however, that Lord Hoffmann's approach is conceptually problematic and is not consistent with corrective justice. This means that just like the *Fairchild* exception, the *Barker* apportionment rule is motivated by considerations of distributive justice and instrumentalism. Given that such considerations cannot provide a coherent foundation for negligence liability, Baroness Hale's open acknowledgment that the solution was based on vague considerations of fairness more accurately reflects the inherent uncertainty created by instrumentalist accounts of negligence.

The developments of the *Fairchild* test therefore serve to illustrate the incoherence that is inevitable when courts adapt the rules of causation to pursue instrumentalist goals rather than corrective justice. Since causation is an essential ingredient of corrective justice-based personal responsibility, there should be no liability when the causal link cannot be proved. While this may leave a pressing social need to compensate victims of a particular disease, as well as a need to deter defendants from engaging in risky conduct, it is not the primary function of negligence law to pursue these objectives. Compensation, deterrence and other instrumentalist goals do not provide a coherent theoretical justification for negligence, so the success of negligence law should not be measured against how well it achieves these goals. As noted in Chapter One,

Weinrib has observed that ‘the legal regime of personal injuries can be organised either correctively or distributively...nothing about personal injury as such consigns it to the domain of a particular form of justice’.¹⁸⁷ In light of the evidentiary gap relating to mesothelioma which prevents attribution of interpersonal responsibility in most cases this is a particular type of personal injury that ought to be addressed through a distributive mechanism.

¹⁸⁷ Ernest Weinrib, ‘Corrective Justice’ (1992) 77 Iowa L Rev 403, 415.

Conclusion

This thesis set out to identify a coherent legal solution to the problem of the ‘evidentiary gap’ in causation in negligence but, as explained in the introduction, it was first necessary to situate the issue within the broader context of causation in negligence more generally. This involved addressing both the nature and function of the tort of negligence as well as the role played by causation within that tort.

In Chapter One, it was argued that the tort of negligence is most coherently conceived as a system of corrective justice. It was seen that Aristotelian corrective justice and Kantian Right together form an account of interpersonal responsibility that prioritises coherence and morality. Causation was shown to be an essential element of interpersonal responsibility, but it was also seen that it has a limited role; causation provides the factual nexus between the claimant and the defendant but it must be supplemented by notions of wrongdoing and damage which sculpt the boundaries of interpersonal responsibility. This must be reflected in the negligence doctrines, so while the various doctrines must join together to ‘articulate a single normative sequence’, the doctrine of causation must have a clearly defined role within that sequence. Chapter One therefore explained the interrelationship between the causation requirement, and the elements of actionable damage, duty of care, breach of duty and quantification of loss. It was also argued that the duty of care is the central concept within the negligence inquiry because it is at the duty stage that the issues of interpersonal responsibility in the relationship between the claimant and defendant are identified, and these issues guide the application of the remaining negligence doctrines in each case.

Having shown in Chapter One that causation has a vital but limited role in interpersonal responsibility, Chapter Two considered the demands that negligence law makes of the causation doctrine and argued that it is essential to isolate factual causation, i.e. the fact of being *a* cause, from evaluative conclusions that a condition was *the* responsible cause. As a test of factual causation, the NESS test was shown to be preferable to the but-for test because it is better

matched to the philosophical account of what it means to be a cause. In straightforward cases the NESS test does not complicate the causal inquiry, and in more complex cases it is able to resolve causal problems that the but-for test cannot. It was also shown that when applying the NESS test it is necessary to clearly identify the damage forming the gist of the action from the outset because the causal problem will vary depending on whether the damage is ‘divisible’ or ‘indivisible’. Applying the NESS test, twinned with a clear understanding of damage, it was possible to clarify the scope and meaning of the *Wardlaw* test of ‘material contribution to harm’. This analysis showed that in cases of divisible damage where there is more than one source of the harmful agent the *Wardlaw* test equates to an application of the but-for test to a portion of the total damage suffered by the claimant. In cases of indivisible damage the *Wardlaw* test may constitute an exception to the but-for test. Crucially, however, it does not constitute an exception to the requirement of causation; it merely compensates for the conceptual inadequacies of the but-for test. The NESS test operates in a simpler way to overcome the same conceptual inadequacies

Having defined clearly the role of causation, and the conceptual requirements for a test of causation, the thesis then turned to evidential problems relating to proof of causation. Chapter Three considered the issue of loss of a chance in medical negligence cases. This argument was raised in *Hotson* and in *Gregg* by claimants who were unable to prove that the defendant doctor’s negligent delay in diagnosis of their illness had caused them to suffer physical harm. Instead they sought proportionate recovery of the physical harm by reformulating the damage as the loss of a chance of avoiding physical harm. Drawing on the understanding of the interrelationship of the doctrines of damage, causation, and quantification previously set out in Chapter One, it was argued that the House of Lords was right to reject their claims because a patient in these circumstances does not have an objective chance of being cured. This means that if proportionate recovery were allowed it would effectively award damages for the physical harm but apply a discount to reflect the degree of doubt surrounding proof of causation. However,

this chapter proposed a more limited role for loss of a chance as an independent form of damage leading to limited recovery in some medical negligence scenarios. It was argued that although an epistemological chance should not normally be considered to constitute damage, it becomes a meaningful loss in this specific context because of the issues of interpersonal responsibility between the doctor and patient in the treatment of existing illness.

Chapter Four turned to the issue of the evidentiary gap that first arose in *McGhee* and subsequently in *Fairchild* and, drawing on the wider understanding of negligence and causation developed in the previous chapters, asked how to achieve a coherent solution to this problem. The problem of the evidentiary gap relates to the scientific understanding of the aetiology of the disease, so the NESS test is unable to resolve the problem but it does enable us to pinpoint the precise nature of the causal problem facing the claimant. This highlighted that it is not possible to define a ‘sufficient set’ of conditions to cause the disease, so it is not possible to draw a meaningful inference of causation. The *McGhee/Fairchild* test of ‘material increase in the risk of harm’ clearly therefore constitutes a relaxation of the causation requirement, so it makes a departure from principles of corrective justice.

This chapter evaluated Lord Hoffmann’s attempt to rationalise the *Fairchild* test in *Barker* by reconceptualising the gist of the action as the ‘risk of harm’ and apportioning liability according to the extent of the defendant’s negligent contribution to the total risk. It was shown that this solution is not consistent with corrective justice, because it conflates the notions of damage and breach by making ‘risk exposure’ central to both.

Both the *Fairchild* test and the *Barker* apportionment principle were shown to be inconsistent with principles of corrective justice and motivated instead by considerations of distributive justice such as the desire to compensate the victims of mesothelioma in *Fairchild*, and the desire to achieve a ‘fair’ balance between the burdens placed on each party in *Barker*. As shown at the outset of this thesis, distributive goals cannot provide a coherent foundation for negligence law; the bipolar structure of negligence liability imposes an artificial limit on the attainment of

distributive goals, and they are mutually truncating. This results in indeterminacy and incoherence in the law. Where the *Fairchild* principle applies it is preferable that it should apply in conjunction with the *Barker* apportionment principle in order to balance the competing interests of claimants and defendants, but this rests on little more than the author's personal preference as, inevitably, any solution premised on distributive principles must. It is therefore important that the application of the *Fairchild* test should be limited to claims for mesothelioma. It is also important that the courts continue to follow advances in medical science and that they revert to orthodox principles of causation as and when the evidentiary gap is closed.

Bibliography

Aagard TS, 'Identifying and Valuing the Injury in Lost Chance Cases' (1998) 96 Mich L Rev 1335

Amirthalingam, K 'The Changing Face of the Gist of Negligence' in Neyers, Chamberlain and Pitel (eds) *Emerging Issues in Tort Law* (Hart Publishing 2007)

Anderson GM, 'Disease Causation and the Extent of Material Contribution' (2006) 15 SLT 87

Atiyah PS, *The Damages Lottery* (Hart Publishing 1997)

Bagshaw R, 'Publication Review: Responsibility and Fault' (2000) 116 LQR 321

Bailey SH, 'Causation in Negligence: What is a Material Contribution?' (2010) 30 LS 167

Barnes DW, 'Too Many Probabilities: Statistical Evidence of Tort Causation' (2001) 64 LCP 191

Beauchamp TL & Rosenberg A, *Hume and the Problem of Causation* (Oxford University Press 1981)

Beever A, 'Cause-in-fact: Two Steps Out of the Mire' (2001) 51 UTLJ 327

— — 'Gregg v Scott and Loss of a Chance' (2005) 24 UQLJ 201

— — *Rediscovering the Law of Negligence* (Hart Publishing 2007)

— — 'Corrective Justice and Personal Responsibility in Tort Law' (2008) 28 OJLS 475

— — 'Formalism in Music and Law' (2011) 61 UTLJ 213

Bevan N, 'Williams v University of Birmingham: Personal Injury – Negligence – Asbestos' [2012] JPI Law 1

Boon A, 'Causation and the Increase of Risk' (1988) 51 MLR 506

Bradford Hill A, 'The environment and disease: association or causation?' (1965) 58 Proc R Soc Med 295

Broadbent A, 'Fact and Law in the Causal Inquiry' (2009) 15 Legal Theory 173

Broadie S & Rowe C, *Aristotle: Nicomachean Ethics: Translation, Introduction and Commentary* (Oxford University Press 2002)

Brown S and Lindsey S, 'Seeing the Wood for the Trees – *Loftus-Brigham* and Apportionment of Damage' (2005) 21 Const LJ 431

Burrows A, 'Uncertainty About Uncertainty: Damages for Loss of a Chance' [2008] JPI Law 31

Cane P, 'Justice and Justifications for Tort Liability' (1982) 2 OJLS 30

— — 'Corrective Justice and Correlativity in Private Law' (1996) 16 OJLS 471

— — *The Anatomy of Tort Law* (Hart Publishing 1997)

— — 'Distributive Justice and Tort Law' [2001] NZ L Rev 401

— — 'Responsibility and Fault: a Relational and Functional Approach to Responsibility' in Peter Cane and John Gardner (eds) *Relating to Responsibility: Essays in Honour of Tony Honore on his 80th Birthday* (Hart Publishing 2001)

— — 'Tort Law as Regulation' (2002) 31 CLWR 305

— — *Responsibility in Law and Morality* (Hart Publishing 2002)

— — 'The Anatomy of Private Law Theory: A 25th Anniversary Essay' (2005) 25 OJLS 203

— — 'Rights in Private Law' in Donal Nolan and Andrew Robertson (eds) *Rights and Private Law* (Hart Publishing 2011)

Caun L, 'Multiple Causes of Injury' [2003] JPI Law 96

Chapman B, 'Ernie's Three Worlds' (2011) 61 UTLJ 179

Coggon D and Taylor AN, 'Causation and Attribution of Disease in Personal Injury Cases: A Scientific Perspective' [2009] JPI Law 12

Coleman J, 'Moral Theories of Torts: Their Scope and Limits: Part 1' (1982) 1 Law and Phil 371

— — 'Corrective Justice and Wrongful Gain' (1982) 11 JLS 421

— — 'Moral Theories of Torts: Their Scope and Limits: Part 2' (1983) 2 Law and Phil 5

— — *Risks and Wrongs* (Oxford University Press 1992)

— — 'The Mixed Conception of Corrective Justice' (1992) 77 Iowa L Rev 427

Davies M, 'The Road From Morocco: *Polemis* Through *Donoghue* to No-Fault' (1982) 45 MLR 534

Englard I, *Corrective and Distributive Justice: From Aristotle to Modern Times* (Oxford University Press 2009)

Finkelstein C, 'Is Risk a Harm?' (2003) 151 U Pa L Rev 963

Fischer DA, 'Tort Recovery for Loss of a Chance' (2001) 36 Wake Forest L Rev 605

— — 'Insufficient Causes' (2006) 94 Ky LJ 277

Fletcher GP, 'Law and Morality: A Kantian Perspective' (1987) 87 Colum L Rev 533

Fumerton R and Kress K, 'Causation and the Law: Preemption, Lawful Sufficiency, and Causal Sufficiency' (2001) 64 LCP 83

Gardner J, 'What is Tort Law For? Part 1: The Place of Corrective Justice' (2011) 30 Law & Phil 1

Gold S, 'Causation in Toxic Torts: Burdens of Proof, Standards of Persuasion, and Statistical Evidence' (1986) 96 Yale LJ 376

Goldberg R, *Causation and Risk in the Law of Torts: Scientific Evidence and Medicinal Product Liability* (Hart Publishing 1999)

— — 'Causation, idiopathic conditions and the limits of epidemiology' (2009) 13 Edin LR 282

Gordley J, 'Tort Law in the Aristotelian Tradition' in David G Owen (ed) *Philosophical Foundations of Tort Law* (Clarendon Press 1995)

Green L, 'The Causal Relation Issue in Negligence Law' (1962) 60 Mich L Rev 543

Green S, 'Winner Takes All' (2004) 120 LQR 566

— — 'The Risk Pricing Principle: A Pragmatic Approach to Causation and Apportionment of Damages' (2005) 4 LP&R 159

— — 'Coherence of Medical Negligence Cases: A Game of Doctors and Purses' (2006) 14 Med LR 1

— — 'Risk Exposure and Negligence' (2006) 122 LQR 386

— — 'Contributing to the Risk of Confusion? Causation in the Court of Appeal' (2009) 125 LQR 44

Grubb A, 'Causation and Medical Negligence' [1988] 47 CLJ 350

Gullifer L, 'One Cause After Another' (2001) 117 LQR 403

Hamer D, "Chance Would be a Fine Thing": Proof of Causation and Quantum in an Unpredictable World' (1999) 23 Melb U L Rev 557

Hart HLA and Honoré T, *Causation in the Law* (2nd ed., Clarendon Press 1985)

Hill T, 'A Lost Chance for Compensation in the Tort of Negligence by the House of Lords' (1991) 54 MLR 511

Hoffmann L, 'Causation' (2005) 121 LQR 592

Hogg M, 'Re-establishing Orthodoxy in the Realm of Causation' (2007) 11 Edin L R 8

— — 'Causation and Apportionment of Damages in Cases of Divisible Injury' (2008) 12 Edin LR 99

— — 'Duties of Care, Causation, and the Implications of *Chester v Afshar*' (2005) 9 Edin LR 156

— — 'Developing Causal Doctrine' in Richard Goldberg (ed) *Perspectives on Causation* (Hart Publishing 2011)

Honoré T, 'The Morality of Tort Law – Questions and Answers' in David G Owen (ed) *Philosophical Foundations of Tort Law* (Oxford University Press 1995)

— — 'Necessary and Sufficient Conditions in Tort Law' in David G Owen (ed.) *Philosophical Foundations of Tort Law* (Oxford University Press 1995)

— — 'Medical non-disclosure, causation and risk: *Chappel v Hart*' (1999) 7 TLJ 1

— — *Responsibility and Fault* (Hart Publishing 1999)

Howarth D, 'Three Forms of Responsibility: On the Relationship Between Tort Law and the Welfare State' (2001) 60 CLJ 553

— — 'Many Duties of Care – Or a Duty of Care? Notes from the Underground' (2006) 26 OJLS 449

Hume, *An Abstract of a Treatise of Human Nature*, ed. JM Keynes and P Sraffa (Cambridge University Press 1938) 11f.

Jansen N, 'The Idea of a Lost Chance' (1999) 19 OJLS 271

— — 'Duties and Rights in Negligence: A Comparative and Historical Perspective on the European Law of Extracontractual Liability' (2004) 24 OJLS 443

Jones MA, 'Proving Causation – Beyond the "But For" Test' (2006) 22 PN 251

— — 'Causation and Psychiatric Damage' (2008) 24 PN 255

Judicial College, *Guidelines for the Assessment of General Damages in Personal Injury Cases* (11th ed., Oxford University Press 2012)

Kant I, *Foundations of the Metaphysics of Morals*, (trans. Lewis White Beck) (Bobbs-Merrill Education Publishing 1969)

Keeler J, 'Thinking Through the Unthinkable: Collective Responsibilities in Personal Injury Law' (2001) 30 CLWR 349

Kelley PJ, 'Causation and Justice: A Comment' (1978) 56 Wash U L Q 635

Khoury L, *Uncertain Causation in Medical Liability* (Hart Publishing 2006)

— — 'Causation and Risk in the Highest Courts of Canada, England and France' (2008) 124 LQR 103

King Jr. JH, 'Causation, Valuation, and Chance in Personal Injury Torts Involving Preexisting Conditions and Future Consequences' (1981) 90 Yale LJ 1353

Kinsky C, 'SAAMCo 10 years on: Causation and Scope of Duty in Professional Negligence Cases' (2006) 22 PN 86

Klimchuk D, 'On the Autonomy of Corrective Justice' (2003) 23 OJLS 49

Kramer A, 'Smoothing the Rough Justice of the *Fairchild* Principle' (2006) 122 LQR 547

Kress K, 'Formalism, Corrective Justice and Tort Law' (1991) 77 Iowa L Rev i

Laleng P, '*Sienkiewicz v Greif* (UK) Ltd and *Willmore v Knowsley Metropolitan Borough Council*: A Material Contribution to Uncertainty?' (2011) 74 MLR 767

Lee J, 'Causation in Negligence: Another Fine Mess' (2008) 24 PN 194

Lewis D, 'Causation' (1973) 70 J Phil 556

Lord Hoffmann, 'Causation' in Richard Goldberg (ed.) *Perspectives on Causation* (Hart Publishing 2011)

Lord Hope of Craighead, 'James McGhee – A Second Mrs Donoghue?' (2003) 62 CLJ 587

Lord Neuberger of Abbotsbury, 'Loss of a Chance and Causation' (2008) 24 PN 206

Lunney M, 'What Price a Chance?' (1995) 15 LS 1

MacCormick N, 'Taking Responsibility Seriously' (2005) 9 Edin LR 168

Mackie JL, *The Cement of the Universe: A Study of Causation* (Clarendon Press 1980)

Martin R, 'Categories of Negligence and Duties of Care: *Caparo* in the House of Lords' (1990) 53 MLR 824

Maskrey S and Edis W, '*Chester v Afshar* and *Gregg v Scott*: Mixed Messages for Lawyers' [2005] JPI Law 205

McBride N, 'Duties of Care – Do They Really Exist?' (2004) 24 OJLS 417

McGregor H, 'Loss of Chance: Where has it Come From and Where is it Going?' (2008) 24 PN 21

McIvor C, *Third Party Liability in Tort* (Hart Publishing 2006)

— — 'Debunking some judicial myths about epidemiology and its relevance to personal injury litigation' (unpublished conference paper, Obligations VI: Challenging Orthodoxy conference, Canada 18th-20th July 2012)

— — 'The Use of Epidemiological Evidence in UK Tort Law' in S Loue (ed.) *Global Perspectives in Forensic Epidemiology* (forthcoming)

Miller C, 'Coal Dust, Causation and Common Sense' (2000) 63 MLR 763

— — 'Judicial approaches to contested causation: *Fairchild v. Glenhaven Funeral Services* in context' (2002) 1 LP&R 119

— — '*Gregg v Scott: Loss of Chance Revisited*' (2005) 4 LP&R 227

— — 'Causation in Personal Injury: Legal or Epidemiological Common Sense?' (2006) 26 LS 544

— — 'Loss of chance in personal injury: a review of recent developments' (2006) 5 LP&R 63

— — 'Liability for Negligently Increased Risk: the Repercussions of *Barker v Corus (UK) Plc*' (2009) 8 LP&R 39

— — 'Causation in personal injury after (and before) *Sienkiewicz*' (2012) 32 LS 396

Moore MS, *Causation and Responsibility: An Essay in Law, Morals, and Metaphysics* (Oxford University Press 2009)

Morgan J, 'Lost Causes in the House of Lords: *Fairchild v Glenhaven Funeral Services*' (2003) 66 MLR 277

— — 'Tort, Insurance and Incoherence' (2004) 67 MLR 384

— — 'The Rise and Fall of the General Duty of Care' (2006) 22 PN 206

— — 'Policy Reasoning in Tort Law: the Courts, the Law Commission and Critics' (2009) 125 LQR 215

Mullany NJ, 'Common Sense Causation – An Australian View' (1992) 12 OJLS 431

Mullender R, 'Negligence Law, the Welfare State, and "Our Moral Life"' (2009) 25 PN 187

Neyers JW, 'The Inconsistencies of Aristotle's Theory of Corrective Justice' (1998) 11 CJLJ 311

Noah L, 'An Inventory of Mathematical Blunders in Applying the Loss-of-a-Chance Doctrine' (2005) 24 Rev Litig 369

Nolan D, 'Causation and the Goals of Tort Law' in Andrew Robertson and Tang Hang Wu (eds.) *The Goals of Private Law* (Hart Publishing 2009)

Oppenheim R, 'The "Mosaic" of Tort Law: the Duty of Care Question' [2003] JPI Law 151

Owen DG, 'Why Philosophy Matters to Tort Law' in David G Owen (ed) *Philosophical Foundations of Tort Law* (Clarendon Press 1995)

Peel E, 'Lost Chances and Proportionate Recovery: *Barker v Corus*' [2006] LMCLQ 289

Perry SR, 'The Moral Foundations of Tort Law' (1992) 77 Iowa L Rev 449

— — 'Professor Weinrib's Formalism: The Not-So-Empty Sepulchre' (1993) 16 Harv J L & Pub Pol'y 597

— — 'Risk, Harm, and Responsibility' in David G. Owen (ed.) *Philosophical Foundations of Tort Law* (Oxford University Press 1995)

— — 'Honoré on Responsibility for Outcomes' in Peter Cane and John Gardner (eds) *Relating to Responsibility: Essays in Honour of Tony Honoré on his 80th Birthday* (Hart Publishing 2001)

— — 'Responsibility for Outcomes, Risk, and the Law of Torts' in Gerald J Postema (ed) *Philosophy and the Law of Torts* (Cambridge University Press 2001)

— — 'The Role of Duty of Care in a Rights-Based Theory of Negligence Law' in Andrew Robertson and Tang Hang Wu (eds) *The Goals of Private Law* (Hart Publishing 2009)

Plowden S and Volpe H, 'Fairchild and Barker in MRSA Cases: Do Fairchild and Barker Provide an Argument for a Relaxation of Causation Principles in Claims for Hospital Acquired MRSA?' [2006] JPI Law 259

Porat A and Stein A, *Tort Liability Under Uncertainty* (Oxford University Press 2002)

— — 'Indeterminate Causation and Apportionment of Damages: An Essay on *Holtby, Allen, and Fairchild*' (2003) 23 OJLS 667

Posner RA, 'The Concept of Corrective Justice in Recent Theories of Tort Law' (1981) 10 JLS 187

Price DPT, 'Causation: The Lords' Lost Chance?' (1989) 38 ICLQ 735

Reece H, 'Losses of Chances in the Law' (1996) 59 MLR 188

Ripstein A and Zipursky BC, ‘Corrective Justice in an Age of Mass Torts’ in Gerald J Postema (ed) *Philosophy and the Law of Torts* (Cambridge University Press 2001)

Robertson A, ‘Constraints on Policy-Based Reasoning in Private Law’ in Andrew Robertson and Tang Hang Wu (eds) *The Goals of Private Law* (Hart Publishing 2009)

— — ‘Justice, Community Welfare and the Duty of Care’ (2011) 127 LQR 370

— — ‘Rights, Pluralism, and the Duty of Care’ in Donal Nolan and Andrew Robertson (eds) *Rights and Private Law* (Hart Publishing 2011)

— — ‘On the Function of the Law of Negligence’ (unpublished conference paper, Obligations VI: Challenging Orthodoxy conference, Canada 18th-20th July 2012)

Robertson DW, ‘Causation in the Restatement (Third) of Torts: Three Arguable Mistakes (2009) 33 Wake Forest L Rev 1007

Robinson GO, ‘Probabilistic Causation and Compensation for Tortious Risk’ (1985) 14 JLS 779

Scherpe J, ‘A New Gist?’ (2006) 65 CLJ 487

Schroeder CH, ‘Corrective Justice and Liability for Increasing Risks’ (1989) 37 UCLA L Rev 439

— — ‘Causation, Compensation, and Moral Responsibility’ in David G Owen (ed) *Philosophical Foundations of Tort Law* (Clarendon Press 1995)

Scott W, ‘Causation in Medico-Legal Practice: A Doctor’s Approach to the ‘Lost Opportunity’ Cases’ (1992) 55 MLR 521

Sheinman H, ‘Tort Law and Corrective Justice’ (2003) 22 Law & Phil 21

Smith J, ‘Causation – the Search for Principle’ [2009] JPI Law 101

Smith SA, ‘Publication Review: Idea of Private Law’ (1996) 112 LQR 363

— — *Contract Theory* (Clarendon Press 2004)

Stanton K, ‘Professional Negligence: Duty of Care Methodology in the Twenty First Century’ (2006) 22 PN 134

Stapleton J, ‘The Gist of Negligence: Part 1 Minimal Actionable Damage’ (1988) 104 LQR 213

— — ‘The Gist of Negligence: Part 2 The Relationship Between “Damage” and Causation’ (1988) 104 LQR 389

— — ‘Duty of Care and Economic Loss: A Wider Agenda’ (1991) 107 LQR 249

— — ‘Tort, Insurance and Ideology’ (1995) 58 MLR 820

- — ‘Negligent Valuers and Falls in the Property Market’ (1997) 113 LQR 1
- — ‘Duty of Care Factors: a Selection from the Judicial Menus’ in Peter Cane and Jane Stapleton (eds) *The Law of Obligations: Essays in Celebration of John Fleming* (Oxford University Press 1998)
- — ‘Publication Review – Causation and Risk in the Law of Torts: Scientific Evidence and Medicinal Product Liability’ (2000) 116 LQR 506
- — ‘Unpacking “Causation”’ in Peter Cane and John Gardner (eds) *Relating to Responsibility: Essays in Honour of Tony Honoré on his 80th Birthday* (Hart Publishing 2001)
- — ‘Lords a’leaping evidentiary gaps’ (2002) 10 TLJ 276
- — ‘Cause in Fact and the Scope of Liability for Consequences’ (2003) 119 LQR 388
- — ‘The Golden Thread as the Heart of Tort Law: Protection of the Vulnerable’ (2003) 24 Aust Bar Rev 135
- — ‘Loss of the Chance of Cure from Cancer’ (2005) 68 MLR 996
- — ‘Two Causal Fictions at the Heart of U.S. Asbestos Doctrine’ (2006) 122 LQR 189
- — ‘Occam’s Razor Reveals an Orthodox Basis for *Chester v Afsbar*’ (2006) 122 LQR 426
- — ‘Choosing What We Mean by “Causation” in the Law’ (2008) 73 Mo L Rev 433
- — ‘Factual Causation and Asbestos Cancers’ (2010) 126 LQR 351
- — ‘The *Fairchild* Doctrine: Arguments on Breach and Materiality’ (2012) 71 CLJ 32
- Stauch M, ‘Causation, Risk, and Loss of Chance in Medical Negligence’ (1997) 17 OJLS 205
- — ‘Taking the Consequences for Failure to Warn of Medical Risks’ (2000) 63 MLR 261
- — ‘Risk and Remoteness of Damage in Negligence’ (2001) 64 MLR 191
- — *The Law of Medical Negligence in England and Germany* (Hart Publishing 2007)
- — ‘“Material contribution” as a response to causal uncertainty: time for a rethink’ (2009) 68 CLJ 27
- Steel S, ‘Uncertainty Over Causal Uncertainty: *Karen Sienkiewicz (Administratrix of the Estate of Enid Costello Deceased) v Greif (UK) Ltd*’ (2010) 73 MLR 646
- — and Ibbetson D, ‘More Grief on Uncertain Causation in Tort’ (2011) 70 CLJ 451
- Steele J, *Risks and Legal Theory* (Hart Publishing 2004)
- — *Tort Law: Texts, Cases and Materials* (2nd ed., Oxford University Press 2011)
- Stevens R, *Torts and Rights* (Oxford University Press 2007)
- Stiggelbout M, ‘The Case of ‘Losses in Any Event’: a Question of Duty, Cause or Damages?’ (2010) 30 LS 558

Thomson J, 'The Raising of Lazarus: The Resurrection of *McGhee v National Coal Board*' (2003) 7 Edin LR 80

— — 'Barker v Corus: Fairchild Chickens Come Home to Roost' (2006) 10 Edin LR 421

Tse MH, 'Tests for Factual Causation: Unravelling the Mystery of Material Contribution, Contribution to Risk, the Robust and Pragmatic Approach and the Inference of Causation' (2008) 16 TLJ 249

Uberoi M, 'Does the Law of Tort Still Recognise The Concept of Individual Responsibility?' (2003) 7 EPIS 1

Voyiakis E, 'The Great Illusion: Tort Law and Exposure to Danger of Physical Harm' (2009) 72 MLR 909

Waldron J, 'Moments of Carelessness and Massive Loss' in David G Owen (ed.) *Philosophical Foundations of Tort Law* (Oxford University Press 1995)

Walker VR, 'Uncertainties in Tort Liability for Uncertainty' (2002) 1 LP&R 175

Weekes R, 'Not Seeing the Wood for the Trees – Risk Analysis as an Alternative to Factual Causation in *Fairchild*' (2003) 12 Nott LJ 18

Weinrib EJ, 'A Step Forward in Factual Causation' (1975) 38 MLR 518

— — 'Toward a Moral Theory of Negligence Law' (1983) 2 Law & Phil 37

— — 'The Insurance Justification and Private Law' (1985) 14 JLS 681

— — 'Law as a Kantian Idea of Reason' (1987) 87 Colum L Rev 472

— — 'Legal Formalism: On the Immanent Rationality of Law' (1988) 97 Yale LJ 949

— — 'The Special Morality of Tort Law' (1989) 34 McGill L J 403

— — 'Corrective Justice' (1992) 77 Iowa L Rev 403

— — 'The Jurisprudence of Legal Formalism' (1993) 16 Harv J L & Pub Pol'y 583

— — 'The Gains and Losses of Corrective Justice' (1994) 44 Duke LJ 277

— — *The Idea of Private Law* (Harvard University Press 1995)

— — 'Corrective Justice in a Nutshell' (2002) 52 UTILJ 349

Williams G, 'Causation in the Law' (1961) 19 CLJ 62

Witting C, 'Duty of Care: An Analytical Approach' (2005) 25 OJLS 33

— — 'The House That Dr Beever Built: Corrective Justice, Principle and the Law of Negligence' (2008) 71 MLR 621

- Wright R, 'Causation in Tort Law' (1985) 73 Cal L Rev 1735
- — 'Actual Causation vs. Probabilistic Linkage: The Bane of Economic Analysis' (1985) 14 JLS 435
- — 'The Efficiency Theory of Causation and Responsibility: Unscientific Formalism and False Semantics' (1987) 63 Chi-Kent L Rev 553
- — 'Causation, Responsibility, Risk, Probability, Naked Statistics, and Proof: Pruning the Bramble Bush by Clarifying the Concepts' (1988) 73 Iowa L Rev 1001
- — 'Substantive Corrective Justice' (1992) 77 Iowa L Rev 625
- — 'Right, Justice, and Tort Law' in David G Owen (ed) *Philosophical Foundations of Tort Law* (Clarendon Press 1995)
- — 'Once More into the Bramble Bush: Duty, Causal Contribution, and the Extent of Legal Responsibility' (2001) 54 Vand L Rev 1071
- — 'The Grounds and Extent of Legal Responsibility' (2003) 40 San Diego L Rev 1425
- — 'Acts and Omissions as Positive and Negative Causes' in Neyers, Chamberlain and Pitel (eds) *Emerging Issues in Tort Law* (Hart Publishing 2007)
- — 'Liability for Possible Wrongs: Causation, Statistical Probability, and the Burden of Proof' (2008) 41 Loy L A L Rev 1295
- — 'Proving Causation: Probability versus Belief' in Richard Goldberg (ed.) *Perspectives on Causation* (Hart Publishing 2011)
- — 'The NESS Account of Natural Causation: A Response to Criticisms' in Richard Goldberg (ed.) *Perspectives on Causation* (Hart Publishing 2011)