In What Circumstances, If Any, Is There a Moral Difference Between Doing and Allowing Harm?

David Harvey Molyneux

Submitted in accordance with the requirements for the degree of Doctor of Philosophy.

The University of Leeds

School of Philosophy, Religion and History of Science.

February 2026

The candidate confirms that the work submitted is her own and the appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

This thesis has only been possible because of the support I have received over many years.

The main source of this support is from my wife, Margaret, who has encouraged and helped me in ways too numerous to mention. Thank you.

More recently, I would like to thank my supervisors, Professor Chris Megone and Dr Christine Nick who have challenged me and helped me to produce this thesis. Their help has been invaluable. All errors and misunderstandings in this thesis are mine, not theirs.

Abstract

This thesis investigates whether there is a moral difference between intentionally doing harm and allowing harm.

My conclusion is that in general, but not always, there is a moral difference between intentionally doing harm and intentionally allowing harm. This moral difference arises in two ways.

Firstly there is a moral difference based on casual factors. Doings of harm, I argue, genuinely cause bad outcomes. By contrast, allowings do not genuinely cause anything, but do feature in a causal explanation of the harm. In general, genuine causation of harm is more blameworthy than causal explanation of harm, and this explains the moral difference between the doing and allowing of harm.

This, however, is not the only source of blameworthiness for the doing and allowing of harm. Another source of moral responsibility is intention; a bad intention can explain (teleologically) the wrongness of the harm, and this is true whether the harm arose from a doing or an allowing.

I argue that talk of explanation, and not causation, is a better way to understand moral responsibility for doing and allowing harm. An account which features *explanation* is more instructive than simply focussing on causation. I show that one instructive way of thinking about explanation is the account offered by Aristotle in his *Physics* and deploy that account in developing my view.

Both causal and teleological explanations contribute to an understanding of the wrongness of doing and allowing harm. Both types of explanation are needed to understand the wrongness of bringing about the harm.

These distinctions have several practical applications. They can help understand the moral difference between euthanasia and withholding/withdrawing treatment, the moral difference between termination of pregnancy and allowing the mother to die, and the moral difference between lying and misleading by omission.

Table of Contents

Acknowledgements **3**

Abstract **4**

CHAPTER ONE - INTRODUCTION ***9***

1.1 Main claim **9**

1.2 Examples **9**

1.3 Why does this distinction matter? **13**

1.4. What is it a difference of? **15**

1.5 Some preliminary definitions **17**

1.5.1 Doings 17
1.5.2 Allowings 17
1.5.3 Acts and Omissions 19
1.5.4 Harm 20

1.6 Other accounts **21**

1.6.1 The sequence account of Phillipa Foot 21
1.6.2 Quinn's Negative/Positive agency account 24
1.6.3 Bennett's 'most of the ways he could have moved' account 25
1.6.4 Woollard's modified sequence account 26
1.6.5 Conclusion from this section 28

1.7 Overview of chapters **30**

1.7.1 Chapter One. Introduction (this chapter) 30
1.7.2 Chapter Two. The Contrast Strategy 30
1.7.3 Chapter Three. Causation, Causal Explanation and Moral Responsibility 33
1.7.4 Chapter Four. Types of Explanation 35
1.7.5. Chapter Five. Intention and Teleological explanation. 36
1.7.6. Chapter Six. Practical applications of this account 39
1.7.7. Chapter Seven. Conclusion 39

CHAPTER TWO – THE CONTRAST STRATEGY ***40***

2.1. Introduction **40**

2.2. What is the contrast strategy? **41**

2.3. Extrinsic and intrinsic moral differences **42**

2.4. How does the contrast strategy work? **46**

2.5. Problems with the contrast strategy **52**

2.5.1 Problems with intuition as a philosophical method 53
2.5.2 Problems with the reliability of contrast cases 55
2.5.3 The intuitive process (whilst being acceptable in general) is not perfect; people can sometimes make epistemic errors. 56
2.5.4 The extrinsic factors in contrast cases are sometimes not made exactly alike in the pair of cases ... 62
2.5.5 The two behaviours (doing and allowing) that are being compared morally are sometimes not adequately differentiated 65

2.6 Provisional Conclusion related to previous three problems **68**

2.7 Problems with the 'Generalisation' move **69**

2.7.1 Rachels's case-based inductive claim 71
2.7.2 Rachels's consistency argument 73

2.8 Against Rachels's generalisation argument **80**

2.9. Some theoretical explanations **84**

2.10 Conclusion **87**

CHAPTER THREE: DOING AND ALLOWING; CAUSATION AND CAUSAL EXPLANATION ***89***

3.1 Introduction **89**

3.2 Causation and causal responsibility **92**

3.3. Moral responsibility for harm **95**

3.4. Prospective v Retrospective Moral Responsibility **95**

3.5. Direct v derived moral responsibility **96**

3.6. Reactive attitudes **96**

3.7 The Standard Claim Regarding Causal Responsibility and Moral Responsibility – some general comments **98**

3.8 Causal and moral responsibility for doings **101**

3.9 Causal responsibility for allowing – allowings do not cause **103**

3.9.1 Argument 1 – causation necessarily involves events – and omissions are not events 104

3.9.2 Argument 2 – Causation requires 'push' (or 'oomph' or 'biff') 108

3.9.3 Argument 3 – Omission causation via counterfactual dependence generates a profligate ontology 109

3.10 Responses to the profligate ontology problem **112**

3.11 The commonsense view that allowings do cause. **120**

3.12 What is causal explanation and how does it differ from causation? **124**

3.13 How can causal explanation be an explanation about causation but not necessarily be causal? **126**

3.14 Can causal explanation help us understand profligate omissions? **132**

3.15 Beebee's salience/adequacy argument **132**

3.16. Problems with the Beebee/Mumford account **135**

3.17 Confusion between causation and causal explanation **137**

3:18 So what evidence is there that there is in fact confusion here? **138**

3.18.1 Livengood and Machery's (L&M) 'The Folk Probably Don't Think What You Think They Think' claim 139

3.19 The moral difference between causation and causal explanation **144**

3:20 Which is a more basic distinction – the doing-allowing distinction or the genuine causation-causal explanation distinction? **146**

3.21 Conclusion **147**

3.21.1 Causal explanation and moral responsibility. 147

3.21.2 Why does it appear that allowings cause? 149

3.21.3 A problem: differentiating between blameworthy and non-blameworthy allowings 149

3.22 A note on terminology **151**

CHAPTER FOUR – OTHER TYPES OF EXPLANATION ***152***

4.1 Introduction **152**

4.2 What is the difference between causation and explanation? **153**

4.3 Causation v explanation in grounding moral responsibility. 157

4.4 Aristotle – explanation not causation – four causes are actually four 'becauses' 162

4.5 Making use of Aristotle's account of explanation 165

4.6 Conclusion 168

CHAPTER FIVE – INTENTION AND TELEOLOGICAL EXPLANATION.....169

5.1. Introduction 169

5.2 What is intention? 173

5.3. Intention matters for the assessment of rightness and wrongness 176

5.4 Intention does not matter for the assessment of rightness and wrongness of actions – Overland's 'all in the mind' argument – and an Anscombian response. 177

5.5. Intention does not matter for the assessment of rightness or wrongness of actions - Rachels' inconsistency argument (and a response from Audi). 184

5.6 Preliminary conclusion 188

5.7. Teleological explanation 191

5.8. Does teleological explanation reduce to causal explanation? 196

5.9. Teleological explanation and moral responsibility? 199

5.10 From teleological explanation to moral responsibility 200

5.11 More about telos 203

5.12 Moral responsibility for doing and allowing harm: coarse distinctions and fine distinctions 204

5.13 The gardener, the Pope and the death of the plants. 205

5.14 Jack and Jill and the blameworthiness of visiting grandma. 206

5.15 The absence in some cases of a moral difference between doing and allowing harm 207

5.16 A note about responsibility for unintentional allowings 208

5.17 In conclusion: the difference between doing and allowing harm. 211

CHAPTER SIX - THE MORAL DIFFERENCE BETWEEN DOING AND ALLOWING HARM – SOME PRACTICAL (MEDICAL) IMPLICATIONS.....213

6.1 Introduction 213

6.2 Ethical issues in end-of-life care 215

6.2.1 Euthanasia can be a doing or an allowing 215
6.2.2 Withholding and withdrawing life-sustaining treatment (LST) counts as an allowing of harm 217
6.2.3 Withholding and withdrawing LST does not cause death 218
6.2.4 Withdrawing LST is not morally different to withholding LST. 223
6.2.5 There is a moral difference between passive euthanasia and active euthanasia 228

6.3 Late termination of Pregnancy 231

6.3.1 Introduction 231
6.3.2 Bennett's argument 232
6.3.3 Responses to Bennett. 237
6.3.4 Problem 1 - Bennett's account of intention 238
6.3.5 Bennett's case comparison is illicit 239
6.3.6 Bennett's account of the distinction between doing and allowing harm is invalid 240

6.3.7 My account of the difference between doing and allowing harm (or between killing and letting die). 242
6.3.8 Conclusion 245

6.4 Lying and Misleading 245
6.4.1 Introduction 245
6.4.2 Lying and other forms of deception; some definitions 247
6.4.3 Arguments that there is no moral difference between lying and non-lying forms of deception 250
6.4.4 Arguments in favour of there being some moral difference between deceiving and non-disclosure (lying is worse than non-disclosure). 253
6.4.5 My account of the moral difference between lying and non-lying deception 257
6.4.6 Objections to this account 258
6.4.7 Conclusion 260

6.5 Chapter Conclusion 261

CHAPTER SEVEN – CONCLUSION 262

7.1 Overall conclusions 262

7.2 Chapter by Chapter review 263

7.3 Final comment 266

CHAPTER EIGHT - BIBLIOGRAPHY 267

CHAPTER ONE - INTRODUCTION

1.1 Main claim

This is a thesis about intentional doing harm and intentional allowing harm – and the presence or absence of a moral difference between these two behaviours.

In this thesis I argue that there is (sometimes) a moral difference between intentional doing and allowing harm. I show that a complete explanation of the moral difference between doing and allowing harm relies on a combination of causal factors and teleological factors.

1.2 Examples

Here are some pairs of examples to set the scene. Firstly, four examples of intentionally *doing* harm.

Intentionally doing harm 1 (modified from Foot 1967, p4)

> **Poisoned grain.** *The people in an African village are close to starvation due to repeated failure of the harvest. Poisoner arranges to send some poisoned grain to the villagers in order to kill the villagers. The villagers eat the poisoned grain. The villagers all die.*

Intentionally doing harm 2 (Rachels (1975, p79)

> **Smith**. *Smith stands to gain a large inheritance if anything should happen to his six-year-old cousin. One evening when the child is taking his bath, Smith sneaks into the bathroom and drowns the child, and then arranges things so that it will look like an accident.*

Intentionally doing harm 3 (Kagan 1988, p7))

> **Push**. Gertrude pushes Bernard into a deep pit hoping that the fall will kill him so that she will inherit the family fortune.

Intentionally doing harm 4 (my case)

> **Crisp packet.** Maurice is late for a meeting and snacks on a packet of crisps whilst taking a short cut across the park. In order to save time, he throws his empty packet on the grass rather than make a detour to the bin at the other side of the park. Maurice thinks – "I'll save time if I don't go to the bin". The park looks slightly less attractive to other park users because of the discarded crisp packet.

These four examples have several things in common. The agents in each scenario *do* something, and the upshot of the doing is that harm occurs. The harm is varied; in these examples it can be death, or possible death, or (more mundanely) a spoiled view.

Now for four examples of intentionally *allowing* harm:

Intentionally allowing harm 1 (modified from Foot 1967p 4)

> **No grain.** The people in an African village are close to starvation due to repeated failure of the harvest. Donor plans to send a sum of money to the village so that they can buy much needed nutritious food. However, at the last

minute, Donor decides to spend the money instead on a new kitchen and does not send the money. The villagers all die.

Intentionally allowing harm 2 (Rachels 1987, p79)

Jones. Jones also stands to gain if anything should happen to his six-year-old cousin. Like Smith, Jones sneaks in, planning to drown the child in his bath. However, just as he enters the bathroom, Jones sees the child slip and hit his head and falls face down in the water. Jones is delighted; he stands by, ready to push the child's head back under if it is necessary, but it is not necessary. With only a little thrashing about the child drowns all by himself, 'accidentally', as Jones watches and does nothing. (Rachels 1975)

Intentionally allowing harm 3 (Kagan 1988, p7)

Failure to warn. Seeing that Bernard is about the fall into a deep pit, Gertrude deliberately refrains from warning him, hoping that the fall will kill him so that she will inherit the family fortune

Intentionally allowing harm 4 (my case)

Left on the grass. Martha is late for a meeting and is taking a short cut across the park. She notices that someone has thrown an empty crisp packet on the grass. She could pick the packet up and take it to the bin at the other side of the park - but leaves it on the grass. Martha thinks – "I'll save time if I don't go to the bin". The park looks slightly less attractive to the other park users because of the discarded crisp packet.

Again, these latter four cases have something in common. All these cases involve the agent in the scenario *allowing* something. Again, the upshot of the allowing is

harm, which seems to be of a similar severity to the harms cited in the doing examples.

Whether doing or allowing, all of the behaviours seen in these cases seem to be morally problematic. So an obvious series of questions arise – are the 'doing' behaviours equally morally problematic when compared with the 'allowing' behaviours? Are the agents in the 'doing' cases as blameworthy as the agents in the 'allowing' cases?

In this thesis I argue that in general (but not always), doing harm is *more* wrong than allowing harm. Agents who do harm are generally (but not always) more blameworthy than agents who allow the same harm.

Why should this be?

My rough explanation is as follows. Doings genuinely *cause* harm, whereas allowings only *causally explain* the harm. Agents are more blameworthy for genuine causation than they are for being part of a causal explanation.

However there is a second way to ground moral responsibility that is applicable for both doings and allowings – this way is via the intention of the agent. An evil intention can increase the blameworthiness of both doing and allowing harm. Sometimes this can make an intentional allowing of harm as blameworthy as an intentional doing of harm.

Obviously this account of the moral responsibility consequent on doing and allowing harm is very rough and ready. I shall expand on it as the thesis proceeds, mainly in Chapters 3 and 5.

1.3 Why does this distinction matter?

Why should the moral distinction (or lack of moral distinction) between doing and allowing harm matter? I suggest that this question is important for two reasons. Firstly, it is important because it impacts on all kinds of practical moral judgements. If it turns out that there is no moral distinction between doing and allowing harm, then this would (for example) speak in favour of active euthanasia being just as morally permissible as passive euthanasia. If it turns out that there is no moral distinction between doing and allowing harm, then this would speak in favour of it being permissible to kill innocent persons in circumstances where it would be permissible to allow innocent persons to die (for example, due to preventable famine or easily treatable infectious diseases). It thus seems important to understand the difference (or lack of difference) between doing and allowing harm in order to make sense of a wide range of practices and moral evaluations.

Generalising from the above, if it turns out there is no moral distinction between doing and allowing harm, then this would suggest that morality is either much *less* demanding than we usually take it to be OR much *more* demanding than we usually take it to be. It might, for example, be morally permissible to kill individuals in order to support and maintain our personal projects - just as we consider it permissible to allow individuals to die, rather than to spend our money on aid projects. Conversely, we might have to provide much more in the way of help to those in need, even to the extent of abandoning our personal projects or suffering a degree of hardship (Barry, Lindauer and Overland 2016, p79).

Secondly, the moral difference between doing and allowing harm is important because of its implications for normative moral theory. In general, act consequentialists believe that all that matters in the moral assessment of a

behaviour is the consequences produced by that behaviour. Whether the consequences are produced by a doing, or by an allowing is, for the act consequentialist, immaterial. For a deontologist or virtue ethicists this is not the case – the *way* that consequences are produced also matters. Thus, if it turned out that there was a moral difference between doing and allowing harm, then one might gravitate towards a deontological ethical theory or a virtue ethics theory. By contrast, if it turned out that there was no moral difference between doing and allowing harm, then one might gravitate towards an act consequentialist moral theory[1].

The topic has attracted attention because of an unresolved controversy within the philosophical literature about the existence of a moral difference between doing and allowing harm. There seem to be two sharply polarised philosophical camps – one camp arguing forcefully that there *is* a moral difference between doing and allowing harm, and one camp arguing equally passionately that this alleged difference is illusory.

Proponents of the '*there is a moral difference'* view include Phillipa Foot (1967, p3), Warren Quinn (1989, p287), Frances Wooll (2008, p17-21) and Fiona Woollard (2016, p7). This moral view is sometimes known as the '*Doctrine of Doing and Allowing'* – or DDA.

Supporters of the '*there is no moral difference'* view include James Rachels (1986, p78), Shelly Kagan (1989, p83-121), and Peter Singer (1979 p181-184). This view has been labelled the *Equivalence theory* (Kamm 1983, p17).

[1] Of course, the difference between consequentialist and deontological moral theories hinges on much more than the difference (or lack of difference) between doing and allowing harm.

Because of these radically differing views, it seems important to develop an account of doing and allowing harm that resolves these differences and can incorporate the insights advanced by these authors about the presence, or absence, of moral difference between the two behaviours.

1.4. What is it a difference of?

I argue in this thesis that there is a moral difference between doing and allowing harm. But what is it a moral difference of? When we assert (as the title of this thesis implies) that there is a moral difference between doing and allowing harm, what are we in fact claiming? There are actually several ways of making this claim.

One obvious way of making this claim is to simply state that that doing harm is morally *worse* than allowing the same harm (a).

Alternatively, one could claim that that duties to avoid doing harm are more *stringent* than duties to avoid allowing harm (b).

Related to this might be the claim that it is *harder to justify* doing harm than to justify allowing harm (c). (Woollard 2015, p13) seems to favour this view.

Alternatively, one might say that the moral responsibility that accrues from doing harm *is greater than* the moral responsibility that accrues from allowing harm (d).

Finally, one might also say that it is *more blameworthy* to do harm than to allow the same harm (e).

One might divide these options into two categories. Some of these distinctions seem to focus on the nature of the behaviour that produces the harm. Options (a), (b) and (c) seem to come into this category. One behaviour may be *worse* than an another, or the constraints against doing this behaviour may be more stringent than the rules

against allowing this behaviour, or one behaviour may be *more difficult to justify* than another.

By contrast, some of these distinctions seem to focus on the nature of the agent who is involved in the doing or allowing. So an agent who does harm may *bear a greater moral responsibility* than an agent who allows harm. Or finally, an agent who does harm may be *more blameworthy* than an agent who merely allows harm[2].

Imagine an agent A who intentionally harms a Victim V. We can claim that doing the harm is morally worse than allowing the same harm (claim a). But this does not translate in a simple mathematical way into a claim about blameworthiness. There are complicating factors. As I will discuss later in this thesis, it may be that that agent does the harm in a non-voluntary way, which will reduce the blameworthiness of the harmful act (perhaps the agent is coerced to bring about the harm). Alternatively, the agent may act without full knowledge of the harm produced by his action or omission. This too may reduce the blameworthiness of the agent.

Throughout this thesis, my focus in mainly on *blameworthiness* – in other words. I will mainly be discussing versions (d) and (e) of the moral difference. I will be arguing that the blameworthiness that accrues from doing is greater than the blameworthiness than accrues from allowing. Put slightly differently, but still carrying the same meaning, I will be arguing that an agent who does harm will be more morally responsible than an agent who merely allows harm.

[2] There is very little written about how these competing concepts connect with each other. One exception to this lacuna in the literature is the work of Chiara Lepora and Robert E Goodin. They argue (2015, p103) that what they call the 'Responsibility Factor' (RF) is related to the wrongness of the wrongdoing (Kc) x Voluntariness of the act (V) x knowledge of the wrongdoing (Kw). Interestingly, Lepora and Goodin equate moral responsibility for a harm with blameworthiness, in much the same way that I do in the discussion above.

But sometimes I will discuss an alternative formulation of the moral difference – perhaps speaking of wrongness or different levels of justification. When I move to one of these alternative formulations, I will make this clear in the text.

1.5 Some preliminary definitions

In order to write sensibly about the topic of this thesis, it important to be clear about some of the terms used. For example, I need to be clear about the meaning of the terms *doing* and *allowing*[3], and what I mean by the word '*harm'*. So before outlining how my overall argument will develop, I want to be clear about the targets of my analysis.

1.5.1 Doings

Doing is usually understood as making something happen – such as the breaking of a window or the killing of a patient (see, for example, Woollard 2015, Chapter 1). This is very much a starting point; I do not want to pre-judge the difference between doings and allowings and will have a great deal more to say about this characterisation of doings later in the thesis.

Here are some day-to-day examples of what are commonly taken to be doings:

- John broke the window by throwing a ball.
- Dr Green killed the patient by giving a lethal injection.
- Elizabeth rolled the stone down the hill
- Gardener Jack killed the dahlias by dousing them with bleach

1.5.2 Allowings

[3] It is generally thought that this distinction is exhaustive and that all types of human conduct that involve harm can be captured by these two terms. Not all philosophers share this view, and some argue that there is a further category of *enabling* harm (see for example, Barry et al, 2014). I'm not going to speak at any length about enabling harm in this thesis; my personal view is that it is more similar to doing than allowing, but nothing hinges in this thesis on this claim.

Allowing, by contrast, is usually understood, (for example, Bach 2013, p51) as *failing to prevent* something from happening when one could have prevented this outcome. Here are some examples of what are commonly taken to be *allowings*:

- Harry allowed the window to break by not catching the ball that was headed for the window.
- Dr Brown allowed the patient to die by not treating his condition (when he could have done).
- Tracey allowed the stone to roll down the hill by not picking it up.
- Gardener Jim allowed the dahlias to die by not watering them.

The common feature in these '*allowing'* examples is that the agent could have prevented the outcome but did not.

Doings and allowings are often intentional, and much (but not all) of this thesis[4] will relate to cases of intentional doings and allowings. All of the examples of doing and allowing mentioned above could easily be construed as intentional doings or allowings. For example, the death of the dahlias, either by administration of bleach or by failure to water could be intentional (perhaps gardener Jim or gardener Jack disliked the owner of the dahlias and intended the death of the plants in order to upset the owner)[5].

However, this is not always the case. There can be *accidental* doings and allowings of harm (accidental doing: the gardener poisoned the plants by choosing the wrong

[4] Much of the literature on doing (and allowing) refers to *intentional* doing and allowing

[5] There are (at least) two ways to think about intentional action. The example of intentional doing allowing given above might be said to employ the '*narrow'* sense of intention where the concept of intention identifies an agent's aims and goals and purposes. The narrow sense of intentional actions is thus in contrast with incidental actions and accidental actions – a described above. There is also a broader sense of intentional, which is the contrary of unintentional. This (broad) sense of intentional action encompasses both the narrow sense of intentional AND incidental actions. I will come back to this topic in Chapter 5 of this thesis, where I will exclusively consider the 'narrow' meaning of the word.

fertiliser; accidental allowing: the gardener was busy with a horticultural emergency and forgot to water the plants) or *incidental* doings and allowings of harm (incidental doing: the oncologist did not intentionally make his patient vomit, but it was an incidental consequence of administering chemotherapy; incidental allowing: the fireman did not intentionally allow the painting to burn, but it was an incidental consequence of saving the owner of the painting).

1.5.3 Acts and Omissions

The distinction between doing and allowing is sometimes conflated with the distinction between acts and omissions. This is an error. There is indeed overlap between the two behaviours – most doings are acts (and vice versa), and most allowings are omissions (and vice-versa). However, the two pairs of concepts can come apart. There can be doings that are omissions and allowings that are acts. Here is an example of an omission that is a doing: I could make a colleague angry by omitting to speak at a meeting (when I had promised to do so). It would be incorrect to say that I merely *allowed* him to be angry: I actually *made* him angry. Similarly, the actor who fails to turn up for a performance is usually thought to actually spoil (a 'doing' word) that performance – he or she does not simply allow it to be spoilt (Foot 1984, p273). One can even *kill* by omission. Clarke (2014, p 186) offers the example of rolling on to an infant and then omitting to move away at the appropriate time.

By contrast, here are some examples of acts that appear to be allowings. A charity donor could allow a child to starve by tearing up (an act) a cheque that the donor planned to give to a charity, and which would have bought vital food supplies for an impoverished community (Foot 1984). Similarly, a doctor can flick a switch (also an act) which will switch off a ventilator and allow a patient to die from lack of oxygen.

In this thesis I am going to concentrate on doings and allowings rather than acts and omissions, as this better captures the difference I am interested in.

1.5.4 Harm

One more clarification. Almost all of this thesis will be about doing and allowing *harm*. This parallels the literature; most discussions of doing and allowing implicitly or explicitly refer to the doing and allowing of *harm*. In terms of this thesis the exact definition of harm is not crucial. The account I have adopted is that of Johannson and Risberg (2022); I see harming as adversely affecting the well-being of an individual (p510). This would restrict the victims of harm to entities whose well-being can be negatively impacted. This would include humans, animals, plants - or groups of these entities. Rather more widely (though not the topic of this thesis) I would think that environments and eco-systems can also be harmed.

For most of this thesis, however, I will be talking about harming *persons*.

One species of harm-doing that features extensively in the literature is *killing* (see, for example, Steinbock and Norcross (1994), which focusses exclusively on that particular species of harm-doing, together with its converse - *letting die)*. This thesis is not particularly about killing and letting die, but I shall refer to both these behaviours in this thesis as part of the discussion of doing and allowing harm[6].

Whilst it is true to say that my main focus will therefore be on the doing and allowing of harm, I have, however, dedicated one chapter of the thesis to exploring the distinction between the doing and allowing of good.

[6] It is worth clarifying that not all killings should be automatically seen as a type of harm – sometimes killing can be considered a benefit to the individual – when for example, the death of an individual causes the relief of unbearable suffering that cannot be relieved in any other way. This is still consistent with my definition of harm, in that the well-being of (say) an individual in unbearable and intractable pain might be improved from strongly negative to zero by ending that individual's life.

Having clarified (at least in a preliminary way) the concepts that I will be using in this thesis, I now go on to outline the main claims and arguments of the thesis.

1.6 Other accounts

I now want to very briefly summarise some other accounts of the alleged difference between doing and allowing harm. This is not an exhaustive survey, and many of the complexities of these accounts have been passed over - it is not my aim to provide a detailed exegesis of the difference (or lack of difference) between doing and allowing harm. What I simply want to do is to outline some other approaches, and to see how my account differs from these accounts.

I want to mention four prominent accounts of the doing allowing distinction.

Three of these accounts (Foot, Quinn, Woollard) argue that there is a moral difference to be found in the doing allowing distinction. By contrast, one account (Bennett) argues that there is no such moral distinction (Bennett refers to the putative distinction as the making-letting distinction).

The four approaches that I present are:

a) The '*sequence'* account of Phillipa Foot,

b) The '*action/inaction'* account of Warren Quinn,

c) The '*modified sequence account'* of Fiona Woollard

d) The '*most of the ways he could have moved'* account of Jonathan Bennett.

1.6.1 The sequence account of Phillipa Foot

Phillipa Foot (1967, 1984) developed her account of the difference between doing and allowing based on the plausible idea of a *harmful* sequence. She claimed that doing harm and allowing harm related in different ways to a harmful sequence.

For example, consider the contrasting cases of *Push*, which is a clear-cut example of a doing of harm, and *Non-interpose*, which is a clear-cut example of allowing harm. (both cases are from Woollard 2015, p3)

First, *Push*.

Push

> *A boulder is in the middle of the road, blocking Bob's route to the hospital. The boulder can only be moved out the way by pushing it down the adjacent slope. Victor is trapped on the slope, directly in the boulder's path. Freeing Victor on detouring around the boulder will delay Bob, so he will not reach the hospital in time. Bob pushes the boulder. The boulder rolls down the slope, hitting Victor and crushing him to death.*

For Foot, in *Push*, Bob does harm because he originates or initiates the sequence that leads to the death of Victor. Here is the sequence broken down into its components:

- Bob pushes the boulder
- The boulder rolls down the hill
- The boulder hits Victor trapped at the bottom of the hill
- Victor dies

It seems plausible to say that Bob does harm because he initiates the sequence that leads to the death of Victor – Bob pushes the boulder.

Now consider a clear-cut case of allowing harm – *non-interpose.*

Non-interpose

The boulder is already rolling toward Victor. Bob could drive his car into the boulder's path, bringing it to a halt. Doing so would delay him so he would not reach the hospital in time. He does not drive into the boulder's path and the boulder hits Victor and crushes him to death.

In this case Bob does not seem to initiate the sequence that leads to the death of Victor. In this sequence the steps can be represented as below:

- An existing sequence is in place
- The boulder rolled down the hill
- *(Bob could interpose his car and stop the boulder, but doesn't)*
- The cart hit Victor trapped at the bottom of the hill
- Victor died

Foot sees *allowing* is forbearing to prevent harm which would be caused by an existing sequence. By this she means that the agent had the ability to prevent a harmful upshot, but does not, thus allowing the harm to materialise.

"....in the case of allowing, a train of events must already have started or be on the horizon; and agent who could stop or prevent it does not do so, and therefore allows it to go on. In the other case, it is he who initiates a sequence, as for instance by shooting someone who dies of the shotgun wound, or pushing into the water someone who then drowns". ((Foot, 1967, p3)

After constructing this dichotomy, Foot is able to link this difference to a moral difference between doing and allowing harm. She does this by arguing that when a harmful sequence is initiated or sustained, the process impinges on a right to non-interference. (Foot calls this a negative right)

By contrast when an agent fails to prevent a harmful sequence from continuing (or, less intuitively enables it to continue) then this seems to impinge on the right to receive aid or receive assistance. Foot thinks that negative rights are less important than positive rights and hence violations of negative rights are less permissible than violations of positive rights.

1.6.2 Quinn's Negative/Positive agency account

Quinn argues (1989) that the difference between doing and allowing (which he called harmful positive and harmful negative agency) could be best described as the difference between harm-producing *action* and harm-producing *inaction.* If the most direct contribution to a harm is an action, then Quinn would call this positive agency. If the most direct contribution is an inaction then Quinn would call this negative agency.

> *...the distinction between harmful positive agency and harmful negative agency would be the distinction between harm occurring because of what the agent does (because of the existence of one of his actions) and harm occurring because of what the agent did not do but might have done (because of the non-instantiation of some kind of action that he might have performed*. (Quinn, 1989, p294)

In the cases of *Push* and *Non-Interpose*, Quinn's account nicely captures the difference between doing and allowing. In '*Push'*, Bob's most direct contribution to the death of Victor is an action - Bob pushed the boulder. Whereas in *Non-interpose*, the most direct contribution to what happened to Victor is an inaction - Bob did nothing.

1.6.3 Bennett's 'most of the ways he could have moved' account

Bennett rejects both the sequence account and the action-inaction distinction. (Bennett, 1995) He feels that both approaches fail to capture the doing/allowing difference (which he calls the making/letting difference). Furthermore, Bennett denies that there can be any moral difference between doing and allowing harm – because his characterisation of the difference between doing and allowing harm does not seem to him to be a reasonable basis for making a moral distinction between the two behaviours.

How does Bennett's argument proceed? Bennett prefers to talk about positive and negative facts, and the difference between positive and negative relevance. Roughly, Bennett sees doing (which he calls making) as having positive relevance to an upshot, and allowing as having negative relevance to an upshot.

Bennett's approach is based on developing an adequate counterfactual theory of causation. He sees that a behaviour is negatively relevant to an upshot when a negative fact about my behaviour is the least informative fact that suffices to complete a causal explanation of it. By contrast, a behaviour is positively relevant to an upshot when a positive fact about my behaviour is the least informative fact about my behaviour that suffices to complete a causal explanation of it.

From this theoretical approach, Bennett develops negative relevance as the condition when most of the ways an agent could have acted at the time *would* have led to the upshot. By contrast, a behaviour has positive relevance if most of the ways the agent could have acted *would not* have led to the upshot.

Here's an example, based on *Push* and *Non-Interpose*. *Push* shows positive relevance as most of the ways the agent could have acted would *not* lead to the cart rolling down the hill (the agent could have read a book, done press-ups, slept etc).

By contrast, in *Non-Interpose*, most of the ways that he could have acted would still lead to the cart crashing over the cliff - the only behaviour that would not lead to this upshot would have been imposing the rock. So the agent could have read a book, done press-ups etc, and the upshot would still have happened.

This is intuitively attractive and nicely explains the difference *Push* and *Non-interpose* and between many doing and allowing behaviours. It reduces the opacity of the doing-allowing difference, whilst paving the way to a claim that the doing allowing distinction cannot be a moral distinction. For how could a moral distinction depend entirely on the difference between the number of different movements that an agent could make?

1.6.4 Woollard's modified sequence account

Woollard's proposal (Woollard, 2015) seems to be a hybrid account - taking elements of Foot's sequence approach, Bennett's '*most of the things he could have done'* account, and (later) Kamm's work on imposition (Kamm, 2007, p21).

So like Bennett, Woollard talks of positive and negative facts about the behaviour of the agent. Like Foot, she thinks of sequences.

Instead of worrying about initiating and sustaining sequences, Woollard's account sees doings as being *part of a sequence*, whereas allowings are *relevant* to a sequence but not actually part of the sequence. Hence she has to explain what being *part of a sequence* means.

Woolard's account of this is exceedingly complex. She begins with what she calls 'relevant facts' - these are "*facts about the agent's conduct on which the upshot depends*". The general idea seems to be a counterfactual one - facts are relevant if they 'make a difference'. If they do not make a difference, then they are not relevant.

Once she has the idea of relevant facts, she then moves on to a sequence. A sequence is a series of relevant facts that lead to an outcome. For example, in *Push*, the sequence is a series of relevant facts as below:

Relevant fact 1 - Bob pushed the boulder

Relevant fact 2 - The boulder rolled down the hill

Relevant fact 3 - The boulder hit Victor

Relevant fact 4 - Victor died.

Woollard then divides relevant facts into substantial facts and non-substantial facts. Relevant substantial facts are suitable to be part of a sequence; relevant non-substantial facts are not suitable to be part of a sequence.

Substantial facts are composed of positive facts together with surprising or non-conventional negative facts.

Negative facts that are not substantial (i.e. negative facts that are not surprising or are conventional) are not suitable to be part of a sequence but can be part of the background conditions for a harm.

Putting this all together, an agent counts as doing harm if and only if a relevant substantial fact about the agent's behaviour is part of the sequence leading to harm. When this occurs, we see the individual as an agent of harm. When there is no relevant substantial fact about the agent's behaviour which is part of the sequence which leads to harm, then he/she is not an agent of harm.

Having established what doing and allowing harm amount to, and how they differ, Woollard now attempts to show that the difference between doing and allowing harm has moral significance. She thus links the first part of her argument (about sequences) to the second part of her argument (about moral significance).

The key idea here is that the moral difference between doing and allowing harm (that it is harder to justify doing harm than to justify allowing harm) protects agents from *harmful impositions*. Harmful impositions (says Woollard) are of two sorts – *causal* impositions whereby an agent imposes on a victim via a *causal* mechanism, and *normative* impositions whereby a potential victim imposes on an agent via a *normative* mechanism. Doing harm to someone involves a causal imposition – Woollard sees this as a chain of substantial facts extending from what the agent does to the victim (or to something that belongs to the victim). By contrast, in allowing harm, the agent does not causally impose on the victim (because there is no chain of substantial facts from agent to victim). Woollard's concern about allowing harm is that unless there is a permission to allow harm, the need to aid a victim normatively imposes on the agent. The agent would be required to use his body or possessions to aid the victim. Woollard sees imposition (both causal and normative) as a bad thing – because both forms of imposition take away something of importance to the agent. She therefore wants to prevent both forms of imposition. She thinks it is possible to prevent causal imposition by imposing a *strong* constraint on doing harm. Similarly, she thinks it is possible to prevent normative imposition by developing a permission to allow harm. If this was not present, then victims would be able to constantly normatively impose on agents so that the agent would have to give up their own resources and possession in order to help others. Hence, Woollard claims, there is an asymmetry between doing and allowing harm.

1.6.5 Conclusion from this section

I have simply described competing accounts of the moral distinction between intentional doing and allowing of harm, without presenting a great amount of detail or

offering a critique of these rival accounts. This will allow a comparison to be made between my accounts and prominent accounts in the literature.

In summary:

a) My account claims that there is a moral difference between doing and allowing harm – unlike Bennett's account which denies this difference

b) My account relies on the difference between genuine causation and causal explanation to explain the moral difference between doing and allowing harm. This differs from Foot's account which relies on the difference between initiation of a sequence and allowing a sequence to continue.

c) My account claims that the important distinction is between doing and allowing. This differs from Quinn's account which relies on the difficult distinction between action and inaction.

d) My account relies on a causal story and a teleological story to explain the moral difference between doing and allowing harm. This is a relatively simple account, in contradistinction to Woollard's account which is complex and requires extra definitions of (for example) substantial facts and imposition.

1.7 Overview of chapters

1.7.1 Chapter One. Introduction (this chapter)

1.7.2 Chapter Two. The Contrast Strategy

In Chapter Two, I claim that, at least some of the time, there *is* a moral difference between doing and allowing harm. In order to make this claim, I need to respond to a line of argument, offered by many (mainly) consequentialist philosophers, which became known as the *Bare Difference Argument*[7] (Perrett 1996, p131). Why is it important to respond to the Bare Difference argument?

There is a strong intuitive case in favour of the idea that there *is* a moral difference between doing and allowing harm. It seems plausible to claim that, across a wide range of real-world examples, doing harm is morally worse than allowing harm – and many have made that claim. So, for example, it is intuitively appealing to say that dropping litter is morally worse than failing to pick up existing litter, or that sending poisoned food aid is morally worse than sending no aid at all, or that throwing stones at windows is morally worse than failing to prevent the same stones from reaching their destination. But this intuitive claim is too quick. What we are really interested in is the moral difference between doing and allowing harm *in itself* – what might be called an *intrinsic* difference. But the real-world moral differences quoted above (litter, food, stones) may not be due to intrinsic differences at all – rather, they may be due to (what might be called) *extrinsic* factors that are differentially associated with doing and allowing harm. Consider for, example, the *likelihood* of the outcome happening. The stone thrown at the window is likely to result in a smashed window;

[7] Sometimes also known as the 'Contrast Strategy' (Kagan 1988, p6)

the failed catch less likely to do so – after all the stone may still miss the window or have insufficient momentum to cause any real damage. Now consider the *cost to avoid* the outcome happening. It costs (usually) very little to avoid dropping litter, whereas to pick up litter may be time-consuming and risky. In both these pairs of examples, the presence of an associated factor (likelihood, cost to avoid, for example) makes it hard to make a moral comparison between the doing *itself* and the allowing *itself*; we may be comparing likelihood and cost in addition to making a pure comparison between doing and allowing harm.

So in order to investigate whether there is an *intrinsic* moral difference between doing and allowing harm, it is necessary to negate (or at least equalise) the influence of these associated factors. Philosophers who deny that there is a moral difference between doing and allowing argue that once these associated factors have been equalised, leaving a 'bare difference' between doing harm and allowing harm – then the alleged moral difference disappears.

Perhaps the most famous example of this argument is the *Bathtub* article offered by James Rachels (1975). Rachels compares two scenarios where a child in a bath dies. In the first scenario, the child is intentionally pushed under the water by an evil cousin ('Smith'), who aims to inherit the child's fortune. In the second scenario, the child accidentally slips under the water and drowns, but an evil cousin ('Jones'), who could easily save the child and actually was planning to push the child under the water intentionally fails to prevent the child from drowning – with the aim of inheriting the fortune.

Rachels' claim is that the behaviours of the two cousins are exactly equal in all ways (intention, motive, probability of success etc) except for the fact that one behaviour (pushing the child under the water) is a *doing* of harm, and one behaviour (failing to

pull the child out of the water) is an *allowing* of harm. Yet it seems that what the 'allowing' cousin did is morally just as evil as what the 'doing' cousin did. Rachels therefore is able to claim that once the associated factors have been equalised, there is *no moral differen*ce between doing and allowing harm. This view is widely shared, and indeed there are plenty of cases in the literature that use similar methodology and reach equivalent conclusions (see for example - Lichtenberg (1982), Bennett (1995)).

Rachels then proceeds to argue that this conclusion can be generalised – he goes on to make the much bolder claim that in *all* comparisons between doing and allowing harm (once the associated factors have been equalised) there is no moral difference between the two behaviours.

I argue that this conclusion is premature, for two reasons. Firstly, it is difficult to completely eliminate associated extrinsic factors without making the cases so complex or under-described that it is impossible to reach a reliable intuitive conclusion about the comparative rightness or wrongness of the doing or the allowing. Secondly, there is considerable variation in the moral conclusions of many of the contrast cases in the literature. For example, the *No-Heroics/Shield* case of Kai Draper (2005) or the *Rescue* case of Phillipa Foot (1984, 1994) suggest an opposite conclusion to that reached in the Rachels case – ie that there is a moral difference between doing harm and allowing harm. This would suggest that the moral conclusion derived from the Bathtub case does *not* generalise, in direct contrast to Rachels' claim.

My provisional conclusion from the analysis of the Contrast strategy is that Rachels' claim that there is *never* a moral difference between doing and allowing harm is suspect. Just because there is no moral difference in the *Bathtub* case does not

mean that there is no moral difference in *all* doing and allowing cases. But such a conclusion, on its own, is only the beginning of an argument. As Alastair Norcross points out (2003, p451) – *'[such a claim] must be accompanied by an explanation of why the distinction is morally relevant in some cases and not in others....'*

The rest of this thesis is therefore constructed around responding to Norcross's challenge.

1.7.3 Chapter Three. Causation, Causal Explanation and Moral Responsibility

I now begin to develop my account of the conceptual and moral distinction between doing and allowing harm.

I differentiate between doing harm and allowing and make the claim that doing harm involves *causing* harm, whereas allowing harm *does not* involve causing harm. This approach is often summarily rejected in the literature, on the grounds that it misunderstands the relationship between causation and allowing. Woollard (2016, Section 4), for example, says this:

> *One natural suggestion is that the agent who does harm causes it to occur; whereas the agent who allows harm doesn't cause it but simply fails to prevent it where she could have done so. This suggestion has immediate moral implications. It seems true by definition (almost) that you can be causally responsible only for upshots that you cause. And it is arguably true that you can be morally responsible only for what you are causally responsible for. This argument gets into trouble when we reflect on the fact that we are often responsible for upshots we allow: the death of the houseplants or the child's illiteracy. When we notice that, in these cases, the plants die or the child remains uneducated because of some failure on the agent's part, it becomes clear that the agent does, in some sense, cause the upshots.*

> *Moreover, most widely accepted contemporary accounts of causation imply that some event or fact involving these agents causes the deaths or illiteracy.*

I argue that this conclusion is too quick. I provide instead a detailed argument that whilst doings cause harm, allowings do not. This difference in causation has the capacity to explain the moral difference between doing and allowing harm – in brief, doing harm is worse than allowing harm. But, as Woollard suggests above, this is not satisfactory – it leaves unexplained the observation that allowings appear to also cause harm – as in the case of the unwatered plants or the uneducated children.

I propose that we can explain this intuition by, following Lewis (1986) and Beebee (2004), differentiating between *causation* and *causal explanation*. I claim that allowings do not cause but do provide a causal explanation for outcomes (in the examples I provide, the outcomes are harms). Causal explanations are explanations, related to causation, but which do not actually involve the process of causation. Doings, by contrast, genuinely cause (and in doing so, of course, also provide a causal explanation). This, I argue, is the difference between doing and allowing. A doing causes something to happen, an allowing merely is part of a relevant casual explanation.

Making use of the concept of causal explanation does seem helpful; it differentiates between doings which both cause and causally explain - and allowings which do not cause but simply provide a causal explanation. However, there is a problem; the reliance on causal explanation does not explain why we blame the gardener for the death of the plants - but do not blame (say) the local mayor whose failure to water also causally explains the death of the plants (had the mayor watered the plants, the plants would not have died). So the difference between causation and causal

explanation, whilst relevant to the difference between doing and allowing, does not by itself provide a convincing explanation of the difference in *moral responsibility* between doing and allowing harm. Consideration of causal explanation is a step along the pathway to understanding the moral difference between doing and allowing harm; but it cannot be the full picture.

1.7.4 Chapter Four. Types of Explanation

I claim in Chapter Four that there is a more comprehensive account of explanation which can help with the problem with the attribution of moral responsibility. To defend this claim, I argue that we ought to develop a richer concept of explanation along broadly Aristotelian lines. Aristotle believed that we only have knowledge of a thing when we have grasped its '*Why'*, or, in other words, its explanation (*Physics* Book II, 3 (2008)). Aristotle argued that explanations could come in four ways or *causes* (Falcon 2023). These are a material cause (*what is it made of?),* as a formal cause (*what is it?),* as an efficient or agent-patient cause (*where does change of motion come from?)* and a final or teleological cause (*what is it's good, or what is it for?).* Not all Aristotelian causes are relevant in every explanation, and, for example, I will not be discussing material and formal causes in the context of doing and allowing. I claim that to fully understand the difference between doing or an allowing, we need not just an explanation involving efficient (agent-patient) causation (what modern philosophy understands as a 'causal explanation' simpliciter) but also a teleological explanation. To offer less than this, is to only have a partial understanding of an action. I suggest that teleological explanations involve the intention with which an agent acts (or fails to act), and hence a full account of moral

responsibility for doing and allowing harm will involve (at least) an account of causal explanation and an account of intention-related explanation.

1.7.5. Chapter Five. Intention and Teleological explanation.

In **Chapter Five**, I flesh out the teleological/intention-related explanation and borrow an account of intention from Elizabeth Anscombe (1963) who sees intentional action as part of a teleological sequence. An agent does A in order to do B, in order to do C in order to do D and so on (the sequence can be short or long). The final step in this sequence must represent some good to the agent (though not necessarily a moral good). This final good allows one to understand and attribute moral responsibility for the preceding steps in the sequence.

Combining an (efficient/agent-patient) causal explanation with a teleological explanation allows us to attribute moral responsibility to a particular doing or allowing. We can thus explain an agent's blameworthiness for intentional *doings* of harm in two ways – by means of a causal explanation (because intentional doing of harm causes the harm*),* and by means of an Anscombean teleological explanation which is part of a teleological sequence (because the harm is intentional and the agent aims at a bad telos). By contrast, for *allowings* of harm, there is no causation of the harm (allowings do not cause) but there is still a causal explanation of the harm, based on the Beebee/Lewis account of casual explanation. And, as in the case of doings, there is a teleological explanation for the responsibility related to the harm allowed - because allowings, like doings, can aim at a bad telos (and are part of a teleological sequence).

The above difference can explain the relative wrongness of doing and allowings of harm. In general, a causal explanation that involves genuine causation (such as we

see in doings) is morally worse than a causal explanation that does not involve actual causation (such as we see in allowing). Thus, causing a broken window is more reprehensible than failing to prevent a broken window, and dropping litter is more reprehensible than failing to pick up litter. The lack of interception of the stone's trajectory, or the lack of litter-picking provides a causal explanation for the window-shattering or the untidiness of the park. In other words, the causal history of the shattering of the window (or the causal history of the messiness of the park did not include an event of the agent blocking the stone's trajectory, or the pedestrian picking up discarded litter. Another way of expressing this causal explanation is this – in a possible nearby world where the agent did catch the stone, or the pedestrian did pick up the litter, then the window may not have been broken, and the park may not have looked untidy. And this seems enough to ground the moral difference; an agent causing a harm seems so much more responsible for the harm than another agent whose role in the harm is a relatively weak counterfactual one.

So much for the causal explanation story; what about the teleological explanation story. How does intention fit in? Anscombe's account (1961, section 5) of intention is such that one can only understand a doing or an allowing by at the same time understanding the intention behind that doing/allowing.

Consider the following pair of examples:

(1) '*John visits his Grandma because he loves and cares for her*'

(2) '*James visits his Grandma because he wants to be included in her will*'.

Despite both actions appearing to be the same – John/James spends the afternoon with his Grandma, sitting by her bedside and presumably cheering her up – the two actions are morally very different. This moral difference cannot be explained in

causal terms, because they both *cause* the same outcome – Grandma feels better. I argue that this difference can only be explained by offering a teleological explanation to supplement the causal explanation.

The same is true of allowings - '*Failing to restart the ventilator because one dislikes the patient'* is a morally different allowing from '*Failing to restart the ventilator because one has made a considered judgement about the patient's quality of life'*. But both behaviours appear to be the same – both involve not depressing the 'On' switch on the ventilator. And both involve the same outcome – the patient dies. I argue that the moral difference between the two behaviours is grounded in the intention underlying the action.

The idea of teleological explanation can help us to understand who can, and who cannot, be held responsible for an allowing in a way that the idea of causal explanation alone cannot. The gardener is responsible for not watering the plants, because it is part of the telos of the gardener to have thriving plants in his care. The local mayor, by contrast, bears no responsibility at all for the demise of the plants, because the thriving of the plants is no part of the mayor's telos (*his* telos is more to do with the thriving of local government).

In Rachels' 'Bathtub' case, both the 'doing' and the 'allowing' cousin have acted equally wrongly even though one cousin has caused the child to die and the other has merely allowed the child to die. One can only understand the equal wrongness of the two cousins' behaviour by considering the role of intention in this case. In other cases, where the intention is perhaps less extreme or even morally neutral,

then the causal difference may play a much greater role in ascribing moral responsibility[8].

1.7.6. Chapter Six. Practical applications of this account

Finally, in Chapter Eight, with this distinction between doing and allowing harm in place, I discuss three practical applications in the field of medical ethics. I consider the moral difference between active euthanasia and passive euthanasia, the moral difference between craniotomy and allowing a mother with obstructed labour to die, and the moral difference between telling lies to patients, and allowing patients to reach their own (erroneous) conclusions. I claim that the moral difference between these pairs of medical strategies can be explained by my account of the moral difference between doing and allowing harm. To take the example of the comparative wrongness of active versus passive euthanasia, active euthanasia is more wrong than passive euthanasia, because active euthanasia (being a doing of harm) actually causes the death of the patient – unlike passive euthanasia which simply provides a causal explanation of the death of the patient.[9]

1.7.7. Chapter Seven. Conclusion

Finally, I conclude in Chapter Eight that my account has greater explanatory power than other accounts and can provide a nuanced explanation of the variable difference between doing and allowing harm, and the difference between the moral responsibility for different sorts of allowings.

---

[9] With the caveat that euthanasia may not always represent a *harm* for the patient

CHAPTER TWO – THE CONTRAST STRATEGY

2.1. Introduction

The main claim that I want to make in this thesis is the claim there is *sometimes (but not always)* a moral difference between intentionally doing and intentionally allowing harm. In order to make good on this claim, I need to respond to an argument advanced by James Rachels which asserts that there is *never* a moral difference between intentional doing and intentional allowing harm. Clearly, if Rachels is correct, then my claim cannot get off the ground.

In order to assert, as I do, that the moral difference is only present sometimes (but not always) I need to respond to Rachel's argument. More specifically, I need to respond to a key part of Rachels' argument, which is the so-called *Contrast strategy*. I argue that the Contrast strategy is methodologically sound, but that the move that Rachels subsequently makes – the so-called *Generalisation move* – is problematic. Because the generalisation move does not go through, this completely undermines Rachels' argument. In other words, there is (or at least may be) a moral difference between doing and allowing harm. This conclusion is important and allows investigation of the causes of the moral difference.

2.2. What is the contrast strategy?

The *Contrast strategy*[10] was initially[11] developed as a method for arguing that there is ***no*** intrinsic[12] moral difference between doing and allowing harm, or between killing and letting die. (i.e., the contrast strategy was developed to show that the Doctrine of Doing and Allowing (DDA) is false).

The contrast strategy was probably first introduced by Michael Tooley (1972) but was popularised by James Rachels with his (now famous) '*Bathtub'* cases three years later (Rachels, 1975). The strategy has been criticised extensively[13], and has also been vigorously defended by a number of philosophers[14]. It is however widely accepted in the literature and quoted extensively in books and reviews about applied ethics and (in particular) about end-of-life care.[15]

In this chapter I argue that with some minimal modification, the Contrast strategy produces reliable moral conclusions about individual pairs of cases. It demonstrates that *at least for some cases*, there does ***not*** appear to be any moral difference between doing and allowing harm. However, importantly, I argue that the conclusion from one pair of cases cannot be generalised to *all* doing/allowing harm cases – so the generalisation move is invalid.

Here is how my argument develops. Firstly (section 3), I outline the difference between intrinsic and extrinsic moral difference. This allows me to (section 4) use this difference to outline the general structure of the contrast strategy, using as an

---

[10] Also known as the '*Bare-difference argument'* or the '*Strategy of fully equalised cases'*
[11] As we will see in the next chapter, the contrast strategy was later used by philosophers to argue *in favour* of the doing-allowing distinction. But that is getting ahead of ourselves - initially (and in most of the literature), the contrast strategy was used to show that there was no intrinsic distinction between doing and allowing
[12] I explain what I mean by an **intrinsic** difference in the next section
[13] For example, by Thomson (1976), Kagan (1988) Kamm (1983), Asscher (2007), Nesbitt (1995), Kopeikin (2019)
[14] For example, by Malm (1992), Perrett (1996), Purves (2011)
[15] For example, in Norcross (2003), Tannsjo (2005)

example, the famous '*Bathtub'* cases of James Rachels. Next, (section 5), I summarise four criticisms of the strategy. These four criticisms are a) epistemic uncertainty, b) unequalised extrinsic factors, c) unclear distinctions between doing and allowing and d) generalisation problem. I argue that these criticisms all throw into doubt the validity of the contrast strategy. However, in section 6, I show that relatively modest methodological changes to the strategy can restore the validity of the contrast strategy for the first three criticisms but not for the fourth criticism.

2.3. Extrinsic and intrinsic moral differences

If there is a moral difference between doing and allowing harm, then this difference could arise in one of two possible ways. The moral difference could be an *extrinsic* moral difference, or an *intrinsic* moral difference (or a combination of the two). In order to understand the contrast strategy, it is important to understand the distinction between these two types of moral difference.

Extrinsic differences between doing and allowing harm, occur because of factors that may (or may not) be associated or correlated with doing and allowing. For example, doing harm is often associated with *intending* harm, whereas allowing harm is often associated with foreseeing harm. Doing harm is often associated with a *malevolent* motive, whereas allowing harm is often associated with a non-malevolent motive or neutral motive. Preventing harm (failing to allow harm) often involves greater *sacrifice* of time and resources than avoiding doing harm. Finally, doing harm is often associated with *certainty* whereas the consequence of allowing harm is often very uncertain. And so on; there actually may be more differences that I have outlined here. The important thing about these extrinsic differences is that they are *usually* associated with doing or allowing harm. But (importantly) they do not have to

be - for example, sometimes allowing harm is both intended and motivated by malice, whereas doing harm may be motivated by indifference and be associated with very uncertain consequences.

Intrinsic differences, by contrast, do not occur due to extrinsic factors. The intrinsic difference between doing and allowing is thus a difference between doing and allowing '*in itself' (or 'in its own right').* Even when all the extrinsic factors are absent or removed (or balance out) an intrinsic difference would still be present.

The doctrine of doing and allowing makes a claim about the *intrinsic* difference between doing and allowing. According to the doctrine of doing and allowing, doing harm is, in itself, (in its own right), morally worse than allowing harm.

It is tempting, when comparing cases, to jump too quickly to a claim about comparative rightness of wrongness of doing and allowing harm. To show how this might happen, I first want to give an example of a moral difference between a pair of cases that *does not* use the contrast strategy. This example will illustrate the problems in drawing a conclusion between pairs of cases when the contrast strategy is not used, and where the difference between intrinsic and extrinsic difference is not made clear.

Imagine two cases, *Famine (1)* and *Famine (2):*

> ***Famine (1).*** *The people in an African village are close to starvation due to repeated failure of the harvest.* ***Smith*** *arranges to send some poisoned grain to the villagers. The villagers eat the poisoned grain. The villagers all die.*
>
> ***Famine (2).*** *The people in an African village are close to starvation due to repeated failure of the harvest.* ***Jones*** *plans to send a sum of money to the village in order to buy much needed nutritious food. However, at the last*

minute, Jones decides to spend the money instead on a new kitchen and does not send the money. The villagers all die.

Most people when asked to assess the relative rightness and wrongness of the behaviours of Smith and Jones would conclude that Smith's behaviour was morally worse than Jones. Indeed, many would argue that Jones's behaviour was not wrong at all. Provided that Jones had earned his money legitimately then he is entitled to donate it or not donate it as he chooses. As Famine (1) involves a doing (sending the poisoned grain), and Famine (2) involves an allowing (omitting to send the nutritious grain), then an initial reading of these two cases might suggest that doing harm is morally worse than allowing harm. It would be wrong, however, to assume that the moral difference between the two cases is purely due to the moral difference between doing and allowing harm. The reason for this is that Famine (1) and Famine (2) are presented in such a way that other relevant moral factors are not equalised between the pair of cases.

Firstly, consider intention. Smith *intended* for the villagers to perish, whereas Jones (at the most) merely *foresaw* that they would die. So any moral difference between the pair of cases may be due to differences in intention, rather than intrinsic differences between doing and allowing.

Now consider motive. Smith was motivated by *malice*, whereas Jones was motivated by eagerness to own a new kitchen. So any moral difference between the pair of cases may be due to differences in motive, rather than intrinsic differences between doing and allowing.

Consider now sacrifice of time and resources. To prevent the villagers' death, Jones would have had to give up his new kitchen, whereas Smith would have to make little

sacrifice if he chose to not send the poisoned food. So any moral difference between the pair of cases may be due to differences in sacrifice, rather than intrinsic differences between doing and allowing.

Finally, consider the certainty or lack of certainty of outcomes. Smith's poison was fairly sure to work, but someone else may have stepped in and sent nutritious food instead of Jones. Thus, any perceived moral difference between the pair of cases may be due to differences in certainty, rather than intrinsic differences between doing and allowing. And so on; there actually may be more differences that I have outlined here. The important thing about these (extrinsic) differences is that they are *usually* associated with doing or allowing harm. But (importantly) they do not have to be - for example, sometimes allowing harm is both intended and motivated by malice, whereas doing harm may be motivated by indifference and be associated with very uncertain consequences.

We can now see that a simple comparison between two cases, one of which is an example of doing harm and one of which is an example of allowing harm is likely to be ineffective in ascertaining whether there is any intrinsic moral difference between doing and allowing harm. There *might* be an intrinsic moral difference between doing and allowing in the *Famine* cases, but this pair of cases would be epistemically inadequate to ascertain this difference. But unless the extrinsic features of the two cases are carefully equalised, then all a comparison of cases (as above) can achieve is a conclusion about the relative *overall* rightness and wrongness of the two cases. This epistemic inadequacy can allow some philosophers (for example, Rachels (1986) and Singer (1993) to argue that the moral difference can be *entirely* explained

by these other moral factors - motive, intention, probability, sacrifice and the rest[16] - and hold that there is no intrinsic difference.

Other philosophers (for example, Foot (1967), Quinn (1989) and Woollard (2015)) accept the existence of extrinsic moral differences between doing and allowing but also believe that there is an intrinsic moral difference between doing and allowing harm - a difference '*in itself*', or '*in its own right*'. For Foot and like-minded philosophers, there is a moral difference between doing and allowing harm that persists even when any extrinsic factors associated with doing or allowing harm and removed or made equal.

I now show how the contrast strategy helps to adjudicate between these opposing views

2.4. How does the contrast strategy work?

The contrast strategy (if it works) aims to demonstrate that there is ***no*** intrinsic difference between doing and allowing harm by in effect removing or balancing the extrinsic factors associated with doing and allowing. The underlying idea is that we construct pairs of cases (known as *contrast* cases) in which one is taken to be a (pre-theoretical) doing, and the other is taken to be a (pre-theoretical) allowing. Importantly, we make all the extrinsic factors identical - in other words we ensure that the *only* difference between the pairs of cases is that one is a doing, and one is an allowing.

It is crucial that the two cases are exactly[17] alike apart from the fact that one is a doing, and one is an allowing.

[16] There could be other factors, but these four are the ones usually cited in the literature

[17] More accurately, '.... the only *relevant* difference between the pairs of cases.....' Clearly irrelevant non-moral details do not necessarily have to be identical - for example in the 'Bathtub' pair of cases, it

> "*It is important that ...cases be exactly alike, except for this one difference...that one involves killing whereas the other involves letting die*". (Rachels (1975) p79)

> *"...we need two parallel situations differing only in that in one a person performs an act resulting in the death of another human being, while in the other she omits to do something, with the same result*". (Singer (1993) p 207)

If this is done, claim the proponents of the contrast strategy, then the moral difference between doing and allowing disappears. This implies that the moral difference between doing and allowing harm is due to one or more extrinsic factors (motive, intention etc) and not due to any putative intrinsic difference between doing and allowing. In other words, whether a behaviour is a doing, or an allowing, makes ***no*** difference in itself.

There is an interesting scientific precedent and analogy to this methodology. It is worthwhile to consider this precedent, as it illustrates what the contrast strategy is attempting to achieve. (It also helps to think about some of the objections to the contrast strategy that I will consider later in this chapter). Consider an experiment to see if the presence of oxygen makes a difference in how various materials burn. This could be achieved by comparing the combustion rate of two candles - one in a bell jar without oxygen and one in a bell-jar with oxygen. Everything else is kept the

---

does not matter for example, that one individual is called 'Smith' and one is called 'Jones' - the names of the different protagonists are different, but this is extremely unlikely to represent a moral difference between the two cases. (Note however that some names might have a moral influence - if one of the evil cousins was called 'Grasper' and the other called 'Gentle' then the names might influence moral assessments of the relative wrongness of the two cases).

same - the type of candle, the size of the bell jar, the presence of other gases, the temperature, the pressure etc. If oxygen makes a difference to combustion, then there should be a difference between the intensity and rate of combustion between the two experiments. If it is not relevant to combustion, then there will be no difference between the intensity and rate of combustion. We do the experiment; clearly oxygen does make a difference; the candle will not burn at all in the absence of oxygen. And the difference that the oxygen makes is an intrinsic difference, a difference *in itself*. Because the other factors are equal, then the difference produced by the presence of oxygen is not due to (for example) it's association with other gases, or with some kind of alteration in the temperature or pressure within the bell-jar.

In the same way, if there is an intrinsic moral difference between doing and allowing harm, then a comparison of two cases, where one is a doing and one is an allowing, and where any extrinsic factors are kept the same (motive, intention etc) - then there should be a moral difference discernible between the two cases. By contrast, if there is no intrinsic moral difference between doing and allowing, then (once extrinsic factors are kept the same) there should be no discernible moral difference between the two cases.

The table below illustrates this clearly:

Table One

	Combustion case	**Doing and allowing case**
Identical factors	Type of candle Size and composition of bell jar Other gases present	Motive Intention Degree of cost/difficulty

	Temperature Pressure	Likelihood of harm occurring
Different factors	Oxygen present or not present	Doing harm or allowing harm
Outcome	Rate of combustion	Overall intuition about rightness or wrongness of the behaviour
Conclusion	Oxygen is an intrinsic factor in the rate of combustion of candle wax	Doing harm is intrinsically morally *worse/better/the same* than allowing harm

Possibly the most well-known example of this strategy is James Rachels' (1975) use of contrast cases which compare the actions of Smith and Jones[18]. Here is Rachel's theoretical justification for this approach:

> *To investigate this issue, two cases may be considered that are exactly alike except that one involves killing whilst one involves letting someone die. Then it can be asked whether this difference makes any difference to the moral assessments. It is important the cases be exactly alike except for this one difference, since otherwise one cannot be confident that it is this difference and not some other that accounts for any variation in assessment of the two cases* (Rachels 1975, p80)[19]

[18] Though Rachels did not coin the term '*Contrast strategy'*. In his initial paper (1975) he did not provide any name for his argument. Later (1986) he seems to have named his argument the '*Bare-difference argument*'. As far as I can see, the term '*Contrast Strategy'* was first introduced by Shelly Kagan in his paper about the additive fallacy. (Kagan, 1988).

[19] Rachels is attempting to show that there is no difference morally between doing and allowing harm (or between killing and letting die) but the lack of difference between any two moral factors of interest could be argued for in the same way

Smith ends the life of his young cousin (to gain a financial reward) by drowning him in his bath. Most people would consider this a doing of harm or a killing. Jones ends the life of his cousin (for the same reason) by not saving the cousin (when he could easily do so) when the cousin slips in the bath. Most people would consider this an allowing or letting die.[20]. Most people who read the case form an intuition that what Smith and Jones did are equally bad. Rachels claims that this fact [that there is no apparent intuitive moral difference between the two cases (despite one behaviour being a doing, and one an allowing)] implies that there is no intrinsic moral difference between doing and allowing, at least in this case[21].

Here are the cases in detail, as described by Rachels:

'Smith and Jones' case (also known as the 'Bathtub' case):

> ***Smith***. *Smith stands to gain a large inheritance if anything should happen to his six-year-old cousin. One evening when the child is taking his bath, Smith sneaks into the bathroom and drowns the child, and then arranges things so that it will look like an accident.*
>
> ***Jones***. *Jones also stands to gain if anything should happen to his six-year-old cousin. Like Smith, Jones sneaks in, planning to drown the child in his bath. However, just as he enters the bathroom, Jones sees the child slip and hit his head and fall face down in the water. Jones is delighted; he stands by,*

[20] Throughout this paper I mainly discuss the distinction between doing and allowing harm. Rachels' discussion, by contrast, concentrates on the distinction between killing and letting die. These differences are obviously not equivalent, as killing and letting die may not always be a harm for the individual concerned. (for an example, consider a patient in unremitting unresponsive pain). Much doing and allowing harm does not lead to death. But, given these caveats, it does not seem unreasonable to talk of the two distinctions in a similar way.

[21] Rachels actually argues from the Smith/Jones case that there is no moral difference between doing and allowing harm *per se*, i.e., in *any* case. This is a much more significant claim and as we shall see later, is not unproblematic.

> *ready to push the child's head back under if it is necessary, but it is not necessary. With only a little thrashing about the child drowns all by himself, 'accidentally', as Jones watches and does nothing. (Rachels 1975)*

Contrast cases need to have the following structure:

Firstly, the two cases should be identical except for the moral factor that is of interest. (in our case, this is whether the behaviour is a doing or an allowing) That is, the extrinsic factors should be exactly the same in both cases. (it may be permissible for there to be differences between the two cases that are clearly irrelevant to the moral question, but for simplicity it is better to make even these the same if possible)

In particular, the extrinsic moral factors should be the same in both cases - for example, the harm produced, the motive, the intention, the likelihood of the bad outcome actually occurring, the ease of avoiding the harmful outcome, etc.

Secondly, the moral factor of interest should be clearly distinguishable between the two cases – for example the behaviour in the 'doing' case should clearly be an example of doing harm and the behaviour in the 'allowing' case should clearly be an example of allowing harm.

Thirdly, there should be a widely agreed intuitive conclusion about the moral comparison of the two behaviours. If the case is a well-constructed case, and if the two behaviours are judged morally equivalent, then it is reasonable to conclude that *in this particular case* there is no moral difference between doing and allowing. By contrast, (again, assuming the case is well-constructed), if there does seem to be a moral difference between the morality of the behaviours then it is reasonable to

conclude that *in this particular case*, there is a moral difference between doing and allowing harm.

At first glance it does seem that the Bathtub case is well-constructed. It does seem that the only difference between the two cases is that one (*Smith*) is a doing, and other (*Jones*) is an allowing.

And, importantly, the extrinsic factors do seem to have been made exactly alike in both the doing and allowing case. Here they are for completeness:

- The motive is the same - to inherit a great deal of money
- The intention is the same - that the child should die
- The consequence is the same - the child does die
- The mechanism of death is the same - drowning
- The commitment from the agent is the same - five minutes (say)
- The likelihood of death is the same - the child will drown

Despite these promising endorsements, there is widespread criticism of the contrast strategy, from a number of angles. Philosophers who are sympathetic to the doctrine of doing and allowing (and some who are antagonistic) argue that the contrast strategy, and this contrast case in particular, is deeply flawed.

I therefore need to look in more detail at these criticisms and ask two questions. Firstly, do the criticisms offered provide undermine the conclusions of the contrast strategy? And secondly, can these criticisms be defused by modification and improvement of the contrast strategy?

2.5. Problems with the contrast strategy

2.5.1 Problems with intuition as a philosophical method

Before commenting about specific problems with the contrast strategy, I want to make a general comment about the principle that *underpins* the contrast strategy; the idea that it is possible to reach meaningful intuitive conclusions about moral cases. The contrast strategy is an intuitive methodology – a thought experiment where we are invited to detect (or fail to detect) moral differences based on our pre-theoretical assessment of the cases presented to us.

What are intuitions in moral philosophy? One helpful definition (Sinnott-Armstrong, Young and Cushman 2010, p246) is that moral intuitions are '*strong, stable, immediate, moral beliefs'*. They are *strong* because they often (but not always) resist evidence to the contrary - though they can be overturned by sufficiently powerful counter evidence. They are *stable* because they are not just temporary whims but persist with time. They are *immediate* because they do not involve steps of conscious reasoning[22].

(It is perhaps helpful at this point to distinguish between intuitions (as I am using this term in this dissertation) and intuitionism. Intuitionism is the theory (espoused historically by Richard Price, HA Prichard and WD Ross, and by such current philosophers as Robert Audi, David McNaughton and Jonathan Dancy – see Stratton-Lake 2002, p1 for an overview) which claims that basic moral propositions are perceived by a special moral faculty, and that moral properties are non-natural properties. My use of the word, 'intuitions' by contrast, does not carry with it any particular metaphysical commitments - it merely asserts that (defeasibly) individuals are able to produce ethical judgements that are strong, stable and immediate.

[22] Another term for intuitive is 'pre-theoretical', which captures the immediacy of intuitive conclusions, together with the emphasis on the absence of steps of conscious reasoning

And, of course, many philosophers have profound reservations about the reliability of intuitions – and there is an extensive literature criticising the value of intuitions in both non-moral (see for example, Kahneman 2011) and in moral reasoning. (see, for example Kagan (1988), Hare (1981), Singer (2005). I don't want to engage in any depth with this substantial body of work here. It is of course possible for individuals to reject the contrast strategy completely, based on a suspicion of any argument based on intuitive conclusions[23]. But it is important to state that if we reject intuition-based reasoning *completely*, then we are also rejecting widely held philosophical beliefs, or beliefs that have stood the test of time, which have been derived from intuitive judgements. The philosophical literature is full of famous arguments that are predicated on intuitive comparisons - arguments such as the Gettier (1963) examples, Judith Thomson's '*Violinist* case (1971), Frank Jackson's '*Mary'* case (1982), and many more. (see Pust (2017) for more examples and references). And if these examples are rejected as being unsound, then it is difficult to know how we can safely rely on in philosophical or ethical reasoning. It seems to be premature therefore to reject such a wide body of philosophical argument purely because we are suspicious of intuition-based reasoning. As Shelly Kagan rightly stated, (Kagan (1988))

> ".... *most contemporary moral philosophy relies heavily on appeals to moral intuitions, and for better or for worse, contrast arguments are no different in this regard*".

So it seems that, even if we accept the limitations (lack of independent calibration, unreliability, lack of explanatory power) of intuition-based judgements, such that the

[23] There are many criticisms of conclusions based on moral intuition. For example (Pust J (2017) has criticised intuitions for being unreliable, or lacking independent calibration, or failing to provide explanation for the conclusions that it supports

contrast strategy depends on, it seems too strong to reject these arguments out of hand. If we did this, arguably, we would threaten all arguments that rely on intuitive moral assessments.

I am therefore going to assume that argument based on intuition, whilst not perfect, is a reasonable starting point for the development of the contrast strategy. More precisely, the limitations of the intuitive approach do not have sufficient power to conclude that the contrast strategy is not fit for purpose in the role of forming conclusions about the relative intrinsic wrongness of harm caused by doing and allowing.

2.5.2 Problems with the reliability of contrast cases

What might be wrong with the contrast strategy? One claim that has been raised is that the strategy does not generate reliable conclusions. So, what factors might jeopardise the reliability of the contrast strategy in identifying whether there is an intrinsic moral difference between a particular pair of cases, one of which is a doing of harm and one of which is an allowing of harm?

It's important to be clear that what follows are criticisms of the contrast strategy, rather than criticisms of the correctness or otherwise of the doctrine of doing and allowing (DDA).

There is an extensive literature on this topic, and I summarise the main themes. There seem to be four main ways that the argument of the contrast strategy can go wrong:

a) The intuitive process (whilst being acceptable in general) is not perfect; people can sometimes make epistemic errors.

b) The extrinsic factors in contrast cases are sometimes not made exactly alike in the pair of cases

c) The two behaviours (doing and allowing) that are being compared morally are sometimes not adequately differentiated

d) The contrast strategy may show that there is (or is not) a moral difference between doing and allowing harm in the particular case chosen, but this moral difference (or lack of moral difference) is not found in all cases of doing and allowing harm.

I will now show that problems (a), (b) and (c) can be remedied and the strength of the contrast strategy can be maintained. However, I argue that problem (d) is intractable and undermines the reliability of the contrast strategy when making claims about all cases of doing and allowing harm.

2.5.3 The intuitive process (whilst being acceptable in general) is not perfect; people can sometimes make epistemic errors.

The worry here is that the reader (assessor) of the contrast case will make an epistemic error - they read the case and perceive that a difference is absent when in fact the difference is present (or vice versa). There are several possible reasons for this to occur. One reason might be that the difference between doing harm and allowing harm is small, and therefore difficult to detect. But this seems improbable – issues around doing and allowing harm and between killing and letting die provoke strong feelings (evidenced by the controversy about active and passive euthanasia),

and it is therefore difficult to imagine that the difference between killing and letting die might be missed.

Another possibility might be that if the background moral features in each case were particularly heinous or objectionable, then these egregious background features might 'swamp' the smaller differences between the moral factors that were being investigated – such as (in the Rachels case) the difference between killing and letting die (this criticism is called the 'swamping' criticism). The '*Bathtub'* case is actually a good example of this rather extreme type of moral scenario – with the inclusion in both scenarios of greed, intention to kill, deceit, and a vulnerable, innocent child. Several commentators have highlighted this potential problem, for example, Kagan (1988) p11:

> *'one or other of the factors that are being held constant is so forceful that the moral status of both cases is extreme – so extreme in fact that the remaining difference between the cases is relatively slight and.....is lost on our insufficiently sensitive intuitions.....both cases may be so horrible, or so wonderful, or even so obviously acceptable – that our failure to notice any difference in the moral status of the cases may not be accepted as sufficient evidence that such a difference is actually lacking"*

A similar point is made by Trammel (1975, p132) in a much more vivid manner.

> *"...the fact that one cannot distinguish the taste of two wines when both are mixed with green persimmon juice does not imply that there is no distinction between the wines*".

Here the underlying idea seems to be that the strong 'flavour' generated by the egregiousness of Smith and Jones's wicked actions can swamp the less intense

difference being doing and allowing harm – and this difference would become apparent if only the masking factors were absent, or less extreme.

Winston Nesbit (1995) makes a similar point whilst reflecting on the importance of intention. Nesbitt argues that the reason that we judge the two bathtub cases to be equally bad is that our judgement is overwhelmed by the *intention* that both Smith and Jones have towards the child. The fact that they both intended to kill an innocent, dependent child (even if Jones ultimately fails in that regard) has the effect of hiding any moral differences between doing and allowing that might have been apparent if the intention was less objectionable.

Another scenario that might generate epistemic uncertainty or error is *complexity*. If the contrast cases that are presented are overly complex, with many subtle distinctions and complications (even if these distinctions and complications were made perfectly alike between the cases) then it might be difficult to come to a clear decision, and even when a decision was reached then this decision might be erroneous.

Linked to the worry about complexity, is a worry about bizarreness, particularly when the writer of the case has taken pains to present a pair of cases where the two cases have made extrinsic factors perfectly alike. Such cases again may be so unusual or so alien to one's experience that it might be extremely difficult to know what one's intuition would be in this situation.

Colombo (2018, p64) makes a related point:

> *".... even though fully equalised cases are actually achieved, they come at the cost of the examples being so contrived and confusing that they compromise intuitions (for instance in the Smith/Jones example)"*

The idea here is that in order to make the extrinsic factors exactly alike (for more on this, see the next section) the cases have to be manipulated, and this manipulation tends to compromise their intelligibility, which in turn leads to epistemic confusion or even error. This may be compounded by the fact that the cases are unusual anyway; how can one reach an opinion when attempting to compare about such bizarre scenarios?

The final epistemic worry that may occur is when the intuitive moral judgement is arrived at too early, or too quickly, or without thinking through all the options. Frances Kamm (2007, p17) argues that this is often the case in contrast cases. For example, in the Smith and Jones case she argues that although our *initial* intuition may be that Smith cases and Jones case are morally equivalent, insufficient time has been spent thinking through whether they are equivalent. For example, if the question was asked, "*What punishment would you be prepared to exact in order to bring the child back to life*" then the equivalence of the moral judgement related to the moral badness of doing might have to be abandoned. So (according to Kamm 2007, p17) the Smith and Jones contrast case might demonstrate the *converse* of the conclusion canvassed by Rachels. And this (says Kamm) is due to a tendency to rush to judgement when presented with unfamiliar cases.

To summarise and briefly respond to this line of criticism. Critics of the contrast strategy seem to have raised legitimate points about epistemic difficulties with the strategy, that might undermine the claim that doing and allowing are intrinsically morally equivalent. Certainly, when creating contrast cases, the philosopher needs to be alert to the epistemic worries related to small differences, 'swamping', complexity and unusualness. But it seems that critics who use these worries to claim that the contrast strategy is an illegitimate indicator of the moral differences

between doing and allowing are being unduly pessimistic. Critics who take this line seem to be unusually pessimistic about the ability of humans to reach reliable intuitive conclusions. However, in Section 5 of this chapter, I will make some modest suggestions about how to make this aspect of the contrast strategy more reliable, and more able to withstand these criticisms.

Response

The theme in the preceding section was that a particular type of case was very likely to generate false or fuzzy intuitions (and hence undermine the cogency of the contrast strategy). These cases involved small differences, swamping, complexity, and unusualness. It is actually helpful to be aware of this; this allows for the design of contrast cases where these sorts of problem features are absent.

To illustrate how cases could be improved, let us concede that the intuitive judgement that the behaviours of Smith and Jones are equally wrong can be challenged. The grounds for challenge might be that intuitive judgements are swamped by strong negative similarities between the two cases. For example, that their intentions are equally odious, that their motives are equally despicable, that their behaviour towards an innocent child are equally callous and deplorable. Furthermore, the whole scenario is rather bizarre and unlikely to occur.

We can use these criticisms to build a better pair of contrast cases. For example, consider a pair of cases which I will call '*Compassionate Mothers'*.

> ***Suzie****. Suzie suffers from a serious wasting disease that is known to progress inexorably and lead to unbearable suffering. One evening when Suzie is taking her bath, Suzie's mother drowns the child, in order to ensure that Suzie will avoid all future suffering. She then arranges things so that it will look like an accident.*

> ***Emma***. *Emma suffers from a serious wasting disease that is known to progress inexorably and lead to unbearable suffering. One evening when Emma is taking her bath, Emma's mother decides to drown the child, in order to ensure that Emma will avoid all future suffering. However, just as she enters the bathroom, Emma slips, hits her head, and falls face-down in the water. Emma's mother stands by, ready to push Emma's head back under the water if it is necessary, but it is not necessary. With only a little thrashing about, Emma drowns all by herself, 'accidentally', as her mother watches and does nothing.*

In this benevolent version of the Smith/Jones case, the extremes of moral viciousness are replaced by a real sense of compassion. Nor is the case as unusual and bizarre as '*Smith and Jones'* - indeed there are many documented examples of a loving parent or carer ending the life of a child or partner in a similar way because of genuine concern about their suffering. Yet in this contrast case, it does seem (as in the original Smith and Jones case) that once extrinsic factors have been equalised, there is no moral difference between the behaviours of the two mothers. (Interestingly, this seems to be true independently of how one feels about the morality of euthanasia). It does therefore seem, by means of imaginative counterexamples, that the epistemic critique of the contrast strategy can be neutralized.

The reliability of the contrast strategy could also be improved by combining groups of contrast cases with differing scenarios. This might be considered a form of triangulation. So for example, if the conclusion about lack of moral difference is the same with '*Bathtub'* and '*Compassionate mothers'* then this would give extra support

to the idea that there really is no moral difference between doing and allowing - at least with this general type of case.

2.5.4 The extrinsic factors in contrast cases are sometimes not made exactly alike in the pair of cases

Recall that the underlying idea of contrast cases is to eliminate the influence of extrinsic factors that may be contingently related to the moral factors that are being compared. An unsatisfactory pair of contrast cases investigating the doing/allowing distinction, for example, would be one where the 'doing' case was associated with a more vicious motive than the 'allowing' case. In order to have confidence that this methodological flaw has been avoided, we need to be sure that any extrinsic features of the case are exactly the same in each case of the pair. Kagan (1988) identified this as a potential problem many years ago – "*a pair of contrast cases is useful only if all other things really are held constant".* Sometimes this can be very difficult to achieve.

There seem to be two sorts of difficulties. The first difficulty relates to problematic detail that is included in the case. For example, in the Smith/Jones pairs of cases, the two evil cousins are not equally successful in their aims. Smith goes into the bathroom with the intention of murdering his cousin and succeeds. Jones goes into the bathroom with the intention of murdering the cousin but fails – all he achieves is to watch the child drown. There appears therefore, to be a clear difference in the *achievement* of intentions between the two cases - Smith achieves his intention, but Jones does not. This probably matters when assessing the validity of the contrast strategy in this pair of cases - the extrinsic factors are certainly not equalised between the two cases. The cases *could* be changed so that Smith has the aim of

killing his cousin, whereas Jones has the aim of letting his cousin die. Now both Smith and Jones achieve their intentions. But this introduces a new extrinsic difference – Smith and Jones now have *different* intentions. This too almost certainly impacts on the validity of the case, as (again) the extrinsic factors are not equalised. The point here is that small alterations to the cases tend to produce new and sometimes unexpected difficulties.

One additional way to solve this problem is to specify the intention of both Smith and Jones using similar language – one might specify, for example, that both Smith and Jones '*aimed for the death of the child by whatever means possible'*. Now they both have the same intention, and both successfully satisfy their aims.

But even this is problematic. An agent who is prepared to aim for the death of a child 'by any means possible' – but who manages to get away with just allowing a death is not typical of an agent who 'just' allows a harm. So again the extrinsic factors are not equalised – we are comparing doing of harm with a rather extreme version of 'allowing' harm.

One response to this unequal extrinsic factor problem would be simply to remove from the case description any problematic extrinsic factors. But this creates a different sort of difficulty – the case will start to appear under-described. Any reader of the case will either be unable to reach any intuitive moral verdict at all – or she will fill in the details for herself and hence create new (implicit) extrinsic differences between the contrast cases. This risks a re-creation of the epistemic problems discussed earlier in this section.

This sort of problem has been highlighted by Asscher (2007) and Hill (2018) and others. Asscher takes a radical view and claims (p 276) that, "*it is impossible to remove all explicit and implicit morally relevant extrinsic differences from examples*".

He sees a tension between detail and ambiguity – too much detail runs the risk that the extrinsic factors will not be exactly the same, too little detail runs the risk that there is not enough material for the reader to make a comparison (or the reader will make a comparison that was not intended by the creator of the case).

To summarise and briefly respond to this line of criticism. The weaker form of this criticism is that it is very difficult to construct pairs of cases where the extrinsic factors in both cases are identical. The stronger form of this criticism is that it is not just difficult but actually impossible to make extrinsic factors identical. Certainly, these types of criticism have merit; unless the extrinsic factors are the same, then the defender of the DDA can simply argue that the absence of moral difference between the two cases is due to differences in doing and allowing 'cancelling out' a difference between non-identical extrinsic factors.

Response

My response to this concern is similar to my response to the epistemic worries raised in sub-section (a). Valid contrast cases may indeed require thoughtful and painstaking construction, but this is not necessarily an argument that suitable cases cannot be devised. I shall flesh out this thought in Section (5) below.

The theme in Section 4b) was that that contrast cases are not reliable because the extrinsic factors are very difficult to equalise or (in Asscher and Hill's version of the criticism) are actually *impossible* to equalise. But this is perhaps unduly pessimistic. It is hard to rule out the possibility that cases could be written that are 'just right'. Such cases should and could be written in such a careful way that the extrinsic factor difficulty can be eliminated. Here is an example.

In the intention problem (*Smith succeeded in his intention, Jones didn't*) discussed in Section 4b, why not describe the intention of both Smith and Jones to be: '*Aiming for the death of the cousin, by whatever means possible'?*

This avoids the difference in *intention* (both evil cousins have the same intention) and also avoids the difference in *intention success* (both evil cousins achieve their intention). This careful approach seems to deal with Asscher's worries about case differences. Alternatively, it may also be possible to safely omit extraneous and confusing detail with little risk of under-describing the case – thus dealing with Asscher's other worry. It therefore seems that the concern about non-equal extrinsic factors is over-stated.

As a final comment about extrinsic factors, if it really is impossible to eliminate the equality of all extrinsic factors, then it may be possible to allow for this inequality in reaching conclusions about relevance or irrelevance. For example, in the '*Bathtub'* case, we could accept the conclusion that there was no moral difference between the two contrast cases but concede that the extrinsic factors in the case are not precisely equalised. If this happened, one could then ask whether the inequality identified is sufficient to jeopardise the conclusion of irrelevance, or whether the conclusion still stands in conditions of less-than-ideal extrinsic factor equivalence.

2.5.5 The two behaviours (doing and allowing) that are being compared morally are sometimes not adequately differentiated

In order to use the contrast strategy to argue that a particular case of doing and allowing demonstrates no moral difference, it is obviously important to make sure that the 'doing' case is really a doing, and the 'allowing' case is clearly an allowing. This criterion actually does seem to be satisfied for the Smith and Jones case

discussed in this chapter - Smith's behaviour is a paradigm of doing and Jones's behaviour is a paradigm of an allowing.

But this nice clear distinction is not always present, and some philosophers have therefore identified this as a flaw in the overall strategy of contrast cases.

For example, Barry, Lindauer and Overland (2014, p71) asked a range of subjects to classify a number of vignettes as examples of either doing or allowing. One case ("*Push*") involved John pushing a cart down a hill and injuring Tom. Another case ("*Stayback*") involved John failing to stop a cart that was already rolling down a hill and subsequently injuring Tom. The respondents were simply asked whether these two cases were cases of doing or allowing. The overwhelming majority of respondents classified '*Push'* as a doing, and '*Stayback'* as an allowing. This suggests, as a minimum, that contrast cases chosen as part as the contrast strategy should be as unequivocal as the cases in this study.

Now consider (again in Barry, Lindauer and Overland (2014, p68), two cases - one called '*Interpose'* and one called '*Remove'*

> ***Interpose****. A cart accidentally starts rolling down a hill. Tom who is sitting at the bottom of the hill won't be injured by the cart if he can get out of the way of its path. John puts a rock on the ground. The rock stops Tom, and he is injured by the cart.*
>
> ***Remove****. A cart is rolling downhill towards a point where there is a rock that would bring it to a stop. John removes the rock. The cart rolls down the hill and injures Tom, who is sitting there.*

In these two vignettes, the respondents disagreed about whether each case was a doing or an allowing, with just over half of those asked classifying '*Interpose'* as

doing harm, and just under half classifying '*Remove'* as doing harm[24]. This emphasises the importance of making sure that the contrast cases are unequivocal - either clearly a doing or clearly an allowing.

A similar criticism could be raised against a contrast case argument utilising a pair of cases, comparing life support being turned off (doing) or being found to be malfunctioning but left unrepaired (allowing). Let's say that this pair of contrast cases are otherwise well constructed, and the reader arrives at the intuition that because of this contrast case, there is no intrinsic moral difference between doing and allowing. An immediate response would be to argue that switching off the ventilator is actually not a doing but an allowing - which undermines the contrast strategy for this pair of cases as now we are comparing two allowings, rather than an allowing and a doing.

To summarise this line of criticism. Barry's (and other's) criticism of the contrast strategy emphasises the importance of clearly demarcated examples of doing or allowing; ambiguity about what counts as a 'doing' and what counts as an 'allowing' will undermine the contrast strategy. But again, I would suggest that achieving this demarcation is not impossible.

Response

The problem in section 4c) can be avoided by being careful in case selection so that doing and allowing are easy to differentiate. In fact, the 'Bathtub' cases seem to achieve this; it is very easy to see that Smith's behaviour is a 'doing' and Jones's behaviour is an 'allowing'. Although there is no totally satisfactory account of the

[24] This particular paper was actually written to support the hypothesis that the genesis of harm is not binary (do, allow) but ternary (do, allow, enable) - but actually provides a good example of the need to be clear on what is being compared in particular contrast cases.

conceptual difference between doing and allowing (See Chapter 4), most people do have a reasonably clear concept of what is involved in doing and allowing, albeit with some confusion on the margins. It seems therefore that the best approach to the design of contrast cases is to carefully avoid ambiguous or intermediate terms - such as '*Interposing'* or '*Removal of barriers'*, or '*Withdrawal/withholding'*. Once this is done, then many of the problems identified in Section 4c can be resolved. Overall, then, with this collection of modifications in place, I believe that this revised contrast strategy will produce sound contrast cases, that are able to generate an argument that (at least sometimes) there is no intrinsic moral difference between doing and allowing.

2.6 Provisional Conclusion related to previous three problems

In this part of the chapter, I have laid out the structure of the contrast strategy and used this structure to classify the criticisms that have been offered to reliability of the strategy.

I have suggested that all the criticisms made about reliability can be classified under three categories. I show that these criticisms all have merit but can be neutralised by appropriate modifications to the methodology of the contrast strategy. In particular, I have argued that the contrast strategy should involve pairs of cases that whenever possible avoid extremes of egregiousness, extremes of complexity and involve scenarios from which it is possible to derive intuitions. I have argued that balancing extrinsic factors between the pair of cases can be difficult, but not impossible. Similarly, I have argued that be clear on what exactly is a '*doing'* and an '*allowing'* can also be difficult but can be achieved. I have given examples of what

might be considered 'improved' contrast cases. I therefore conclude that the contrast strategy is a valuable method of criticising the doctrine of doing and allowing.

In the next section I will consider whether a difference (or lack of a difference) between doing and allowing might be generalised to all cases of doing and allowing. But even if it turns out that the insights from the contrast strategy cannot be generalised to all cases of doing and allowing, the limited conclusions derived from this chapter are considerable. For if *just one* case shows there is no intrinsic moral difference between doing and allowing, then we can conclude that the doctrine of doing and allowing is not universally true. In other words, there is *not always* an intrinsic moral difference between doing and allowing harm.

2.7 Problems with the 'Generalisation' move

So far, I have argued that the Contrast strategy was able to demonstrate, at least for the pair of cases I discussed, (the 'Bathtub' pair of cases), that there is no intrinsic moral difference between doing harm and allowing harm with the same intention. Here's a reminder of that case:

'Bathtub' case (Rachels 1975, p79):

(1) ***Smith***. *Smith stands to gain a large inheritance if anything should happen to his six-year-old cousin. One evening when the child is taking his bath, Smith sneaks into the bathroom and drowns the child, and then arranges things so that it will look like an accident.*

(2) ***Jones***. *Jones also stands to gain if anything should happen to his six-year-old cousin. Like Smith, Jones sneaks in, planning to drown the child in his*

> *bath. However, just as he enters the bathroom, Jones sees the child slip and hit his head and fall face down in the water. Jones is delighted; he stands by, ready to push the child's head back under if it is necessary, but it is not necessary. With only a little thrashing about the child drowns all by himself, 'accidentally', as Jones watches and does nothing. (Rachels 1975)*

Let's be clear - at this stage, all the argument shows is that *at least for this one pair of cases (the Bathtub pair)*, there is no *intrinsic* moral difference between intentionally doing and intentionally allowing harm. But Rachels is not satisfied with this limited conclusion; he wants to say something more - that in *all* pairs of cases, where one is an example of doing harm and the other is an example of allowing harm, then there is no intrinsic moral difference between doing and allowing harm. (of course there may be *extrinsic* moral difference; but it is the role of the contrast strategy to eliminate extrinsic moral differences between the two cases of the case pair)

Rachels is very confident about this conclusion; he says (Rachels 1986 p 111) - "*whatever reasons there may be for judging one act to be worse than another, the bare fact that one is a killing*[25] *while the other is letting die, is not among them*".

I plan (for convenience) to call this claim the '*Generalisation claim'*. Kagan (1988, p12) calls this claim the *Ubiquity thesis.* This latter expression require some explanation - 'generalisation claim' has the advantage that the term is fairly self-explanatory.

[25] Rachels sometimes talks about the difference between doing and allowing harm, and sometimes about the difference between killing and letting die - but he is talking about the same sort of difference each time. In this essay I have tended to stick to the doing harm/allowing harm difference.

In the first part of this essay I outline Rachels's two main arguments for his conclusion - a (fairly weak) argument based on a number of cases similar to the 'Bathtub' case that support Rachels's conclusion, and a more theoretical argument based on consistency. In the second part of this section I challenge Rachels's two arguments. I show that there are pairs of doing/allowing cases that reach an exactly opposite conclusion than the one reached by Rachels in the 'Bathtub' case. And secondly, I show that Rachel's theoretical argument from consistency does not work; generalisation is only possible for *some* pairs of cases.

I thus argue that Rachels is wrong; the conclusion from the '*Bathtub'* case cannot be generalised to other examples of doing and allowing harm. In some of the 'contrast' cases there is an intrinsic moral difference between doing and allowing harm; in other contrast cases there isn't.

2.7.1 Rachels's case-based inductive claim

How does Rachels go about arguing for his conclusion? Initially he simply *assumes* the truth of the generalisation claim, without offering any particular argument. (the initial 'Bathtub' paper of 1975 does not talk about generalisation at all). An example of Rachels's assumption is the claim that if there is no intrinsic moral difference between doing and allowing harm in the '*Bathtub'* cases, there will also be no intrinsic moral difference between doing and allowing harm in euthanasia/removal of treatment cases. In 'The End of Life' (1986 p113), for example Rachels says this:

".... the point will be the same even in those cases: [doctors ending a patient's life for reasons of compassion] the difference between killing and letting die does not ***in***

***itself**, make a difference, from the point of view of morality.....if the doctor's decision was the right one, then the method he used is not itself important*"

Rachels does not provide any evidence for this extrapolation.

Secondly, (even though Rachels does not marshal these cases towards his overall conclusion) there are many cases in the literature which gives support to the conclusion reached by Rachels in his '*Bathtub'* case. The cases have the same structural features as the 'Bathtub' case. These cases are numerous, and I do not plan to discuss them in any detail in this essay. They are listed in the footnotes below[26][27][28][29][30], and it is relatively easy to construct additional, similar, cases. It is

---

[26] **Jones and Smith and the bomb (Tooley 1972, p108)**

(1) *Jones sees that Smith will be killed by a bomb unless he warns him. Jones's reaction is: "How lucky, it will save me the trouble of killing Smith myself". So Jones allows Smith to be killed by the bomb even though he could easily have warned him*

(2) *Jones wants Smith dead, and therefore shoots him.*

[27] **Gertrude and Bernard and the deep pit (Kagan 1988, p7)**

(1) *Gertrude pushes Bertrand into a deep pit, hoping that the fall will kill him so that she will inherit the family fortune.*

(2) *Seeing that Bertrand is about to fall into a deep pit, Gertrude deliberately refrains from warning him, hoping that the fall will kill him so that she will inherit the family fortune.*

[28] **Gertrude and Bernard and the deep pit - self defence version (Kagan 1988, p7)**

(1) *In order to defend myself against the aggressor, I push him into a pit, expecting the fall to kill him.*

(2) *In order to defend myself against the aggressor, I refrain from warning him about the pit into which he is about to fall, and I expect the call to kill him.*

[29] **Two sons and the poison. (Tooley (1973, p104)**

Two sons are looking forward to the death of their nasty but very wealthy father. Tired of waiting, they decide, independently of one another, to kill their father.

(1) *The first son puts some poison in his father's whiskey and is discovered doing so by his brother who was just about to do the same thing.*

(2) *The second son then allows his father to imbibe the deadly drink and refrains from administering an antidote which he happens to have.*

[30] ***Two buttons (Lichtenberg 1982, p19)***

Imagine a machine which contains two children John and Mary.

(1) *If one pushes a button, John will be killed, but Mary will emerge unharmed.*

(2) *If one does not push the button, John will emerge unharmed, but Mary will be killed.*

worth pointing out that part of Rachels's argument is after all an inductive argument, and the strength of the generalisation argument is therefore weakly related to the number of cases offered in which there does not seem to be any intrinsic moral difference between doing and allowing harm[31].

2.7.2 Rachels's consistency argument

In later work, particularly in his book, 'The End of Life' (1986) and his article 'Reasoning about killing and letting die' (1981), Rachels presents a theoretical argument for his claim that the 'Bathtub' argument can be generalised.

Here (1986 p123) Rachels is defending his argument against the criticism that he has rushed too quickly to a generalisation claim:

> *".....it is suggested that the bare [contrast] argument commits an elementary fallacy - the fallacy of hasty generalisation. It leaps from one example [the 'Bathtub' case], in which the distinction appears to be unimportant to the general conclusion that the distinction is* ***never*** *(*my emphasis) *important. But why should this be.....perhaps the truth is simply that the difference between killing and letting die is sometimes morally important and sometimes not....."*

Rachels calls this intermediate claim ('*sometimes morally important and sometimes not'*), the *compromise* view. Rachels strongly rejects the compromise view. He goes on to develop arguments based on *consistency* to support this view. The basic idea is that moral facts do not operate in that way - moral facts cannot sometimes '*count as a good reason to support a moral judgement and sometimes not*". (Rachels 1981 p468).

[31] Assuming of course that the various case are not just the same case repeated in almost identical contexts

Rachels develops his argument by the use of several examples.

In '*Frank and John'*, both Frank and John are equally stingy and equally dishonest. Rachels supposes that it would be inconsistent to assert that '*although stinginess and dishonesty count against John, the same does not apply to Frank'*. (Rachels 1981 p468). If the reason that Frank is a bad man is his stinginess and his dishonesty, then if John is equally stingy and dishonest, he must be equally bad (assuming other factors are equal). The dishonesty reason and the stinginess reason cannot suddenly lose their moral strength or change the direction of their moral impact.

In '*Brown and Black'*, Rachels develops a similar consistency argument. Imagine that '*Brown'* and '*Black'* are both candidates for (an American) local election. They have identical views on gun control and affirmative action (and there are no other differences in their political beliefs or capabilities). Rachels claims that it would be inconsistent to say that '*Brown's position on gun control or affirmative action goes on the list in her favour, but that Black's position doesn't even go on the list'*. (Rachels 1986 p125) This is a similar argument to the Frank/John argument - normative reasons do not seem to be able to change their influence on a normative judgement when simply moved from one person to another or from one situation to another.

Both these examples seem plausible – appeals to consistency certainly have rational appeal.

Rachels now moves from this claim (that moral principles such as honesty and stinginess have a consistent moral import when moved from person to person or from situation to situation) to a parallel claim about the consistency of moral principles such as the wrongness of harm through doing and the wrongness of harm through allowing. If there is no difference between harm achieved by doing and

harm achieved by allowing in the *Bathtub* case, then there will be no difference in *any* situation where harm is produced by a doing and harm is produced by an allowing.

Rachels analogy is easy to see. If there is no difference in the degree of stinginess or the degree of dishonesty between these two individuals, then there can be no difference in the assessment of their character. By analogy with the wrongness of the case of doing and allowing, if the fact that Smith's act is a doing of harm, and Jones's act is an allowing of harm is a reason for judging 'Smith's act to be morally equivalent to 'Jones' act, then this is a reason for judging *any* pair of cases to be morally equivalent, where one is a doing of harm and the other is an allowing of harm. (again, with all other morally relevant factors being equal between the two cases)

This then is Rachels' argument. In the next sub-section, I will attempt to show that this argument is false and suggest what the implications of this conclusion might be.

Response

I now want to present a counterargument and to argue that Rachels' conclusion is false.

Firstly, I present a case-pair that generates an opposite intuition to the intuition seen in the 'Bathtub' case. I show why this a problem for Rachels.

Secondly, I show that Rachels' generalisation argument, whilst plausible, does not get Rachels to where he wants to be.

a) Contrasting intuitions

Here is a case pair (Foot, 1994, p283), which seems structurally the same as the 'Bathtub' case, but which produces an opposite conclusion.

Rescue I. We are hurrying in our jeep to save some people - let there be five of them - who are imminently threatened by the ocean tide. We have not a moment to spare, so when we hear of a single person who also needs rescuing from some other disaster, we say regretfully that we cannot rescue him, but must leave him to die.

Rescue II. We are again hurrying to the place where the tide is coming in, in order to rescue the party of people but this time it is relevant that the road is narrow and rocky. In this version the lone individual is trapped (do not ask me how) on the path. If we are to rescue the five, we would have to drive over him. If we stop, he will be all right eventually, he is in no danger unless from us. But of course, all five of the others will be drowned.

Rescue 1 is clearly a case of allowing the single trapped person to die. *Rescue 2* is clearly a case of killing the person on the road. There don't seem to be any confounding factors. In both scenarios, the intention is the same - to '*save a trapped person'*. Yet equally clearly it seems that there is a moral difference between doing and allowing - what is done in *Rescue 2* seems worse than what is done in *Rescue 1* Recall that Rachels's contrast argument was an inductive one, so gained plausibility by being able to offer a number of cases that reached the same conclusion. As with Rachels' argument, this argument is an inductive one and will gain extra plausibility by increasing the number of cases that reach the same intuitive conclusion. So, as I did in the first section, I give examples of these in the footnotes below. The common factor between these cases is that they all involve the plausible intuition (in

contrast to '*Bathtub'*) that there is an intrinsic moral difference between doing and allowing harm. [32,33,34,35,]

If this conclusion from cases is right, then this causes enormous problems for Rachel's generalisation argument. Rachels wants to argue from the Bathtub case that ***all*** contrast cases will produce the same result - there is no difference between

---

[32] **Push and Non-Interpose (Woollard 2015, p3)**

Push: A boulder is in the middle of the road, blocking Bob's route to the hospital. The boulder can only be moved out of the way by pushing it down the adjacent slope. Victor is trapped on the slope, directly in the boulder's path. Freeing Victor or detouring around the boulder will delay Bob so he will not reach the hospital in time. Bob pushes the boulder. The boulder rolls down the slope, hitting Victor and crushing him to death.

Non-interpose: The boulder is already rolling towards Victor. Bob could drive his car into the boulder's path, bringing it to a halt. Doing so would delay him so he would not reach the hospital in time. He does not drive into the boulders path and the boulder hits Victor.

[33] **Trevor and Zager (Kopeikin 2019, p559)**:
Trevor. Trevor is steering his new boat through a narrow channel. On both sides of him are shallow, rocky waters that would cause irreparable damage to his boat were he to steer into them (he would make it over the rocks but his engine would be destroyed, his hull would be compromised, he would get banged up a bit, and he would have to wait for an hour before being rescued). It's windy, making the lake choppy. Trevor has to fight the water to keep his boat going straight. As Trevor navigates the channel, he sees a swimmer in front of him that he'll run over and kill if Trevor continues in the channel. Trevor does everything he can to alert the swimmer but the swimmer doesn't respond. Though Trevor would prefer not killing the swimmer, he strongly doesn't want to destroy his boat and be physically injured. Trevor continues to fight the choppy waters and pilot down the channel, running over and killing the swimmer.
Zager: Zager also just bought a new boat, is piloting it through a narrow channel, etc. The only difference is that there is no swimmer in front of him. Rather, he sees a drowning swimmer on the other side of the shallow, rocky waters to his left. Zager could save the swimmer, but it would destroy his boat (the same way that it would destroy Trevor's boat) and would hurt him the same way. Not wanting to destroy his boat, Zager continues down the narrow channel and lets the swimmer drown.

[35] **Shooting and the charity donation. (Trammell R 1975, p 131)**

If someone threatened to steal $1000 from a person if he did not take a gun and shoot a stranger between the eyes, it would be very wrong of him to kill the stranger to save his $1000. (Killing A to save $1000)

But if someone asked from that person $1000 to save a stranger, it would seem that his obligation to grant this request would not be as great as his obligation to refuse the first demand - even if he had good reason for believing that without his $1000 the stranger would certainly die. (Allowing A to die to save $1000)

doing and allowing harm. But here is at least one case-pair – Rescue I and II (and quite probably several more), where this is clearly *not* true - in these cases there *does* seem to be an intrinsic moral difference between doing and allowing harm, despite all other factors being equal.

I can find no evidence that Rachels (or indeed any of those philosophers who claim there is no intrinsic difference between doing and allowing harm) have responded in any way at all to these many case pairs examples that purport to show a difference between doing and allowing harm. Rachels simply directs the reader to his theoretical argument and sticks to his 'generalisation' claim. But if Rachels did respond, how could he neutralise the importance of cases like '*Rescue I and Rescue II*?

It may be argued that there is something wrong with this latter group of cases. If this was true, then this would remove the challenge to Rachels's claim about generalisability. Recall that we were assessing the validity of the original contrast cases, and we listed several desiderata for them to be good cases.

As a reminder, they are: a) They generate clear and reliable intuitive assessment of the rightness or wrongness of the case, b) All imaginable extrinsic factors are equal in the two cases, c) There is a clear demarcation between doing and allowing harm. Are these factors satisfied for *Rescue I* and *Rescue II*? I would argue that they are. The intuition that *Rescue I is* morally less troublesome than *Rescue II* seems well-founded. The extrinsic factors seem equal in both cases - in both cases there is a threat, in both cases the motive is to prevent others being killed, in both cases there is no intention that Jones is killed. And it seems (to me at least) that the behaviour in '*Rescue I'* is an allowing and in '*Rescue II'* it is a doing. So, it does seem that this case meets the desiderata discussed earlier.

More seriously, perhaps the whole contrast strategy is flawed? In the previous chapter, we discussed (and rejected) the worry that the methodology, based as it is on an intuitive (pre-theoretical) moral judgement, was unsatisfactory. Perhaps we were too quick to reject this worry. Perhaps the reliance on pre-reflective judgements makes the whole contrast strategy suspect? This is still possible but (as discussed before) a reluctance to base conclusions on pre-reflective judgements would undermine much of philosophy. In any case, if there are problems with this case, there will also be problems with the whole of the Contrast cases discussed in part one of this chapter[36].

If the arguments in the preceding paragraphs are correct, then this creates a number of problems for Rachels.

Firstly, it is no longer true that there is no intrinsic moral difference between *every* pair of doing and allowing cases. There is at least one case pair - '*Rescue I and Rescue II'* (and maybe more) where it appears that (contrary to the Contrast strategy) there *is* an intrinsic moral difference between doing and allowing harm. Note that it only requires *one* case to disprove a hypothesis. The presence of the 'Rescue I and Rescue II' case pair thus disproves the claim offered by Rachels in the previous chapter.

Secondly, this conclusion, if correct, puts pressure on Rachel's generalisation argument - it looks like there is something wrong with it.

So let's look at the generalisation argument again.

[36] Although I will cover this in the next essay it is worth making a brief point here about intention. There is an argument that in the 'Bathtub' case, the behaviours of Smith and Jones were morally equivalent simply because they were both intentional - and it is this intentionality that helps us to reach the conclusion that the two cases are morally equivalent.

2.8 Against Rachels's generalisation argument

Rachels's claim (see Section 1 of this essay) is that rules of consistency imply that if a moral factor makes (or fails to make) a difference in one case, then it must make a difference in all relevantly similar cases. This is what allowed Rachels to argue that because there is no intrinsic moral difference between doing and allowing harm in the '*Bathtub'* case, then there is no difference between doing and allowing in any pair of cases where one is a doing of harm, and one is an allowing of harm (and there is no other relevantly relevant fact). And this seems, at first sight, very plausible. How could it be that the difference between doing and allowing harm does not matter in *this* case but does matter in *that* case. In the same way, it is difficult to see that the fact of being dishonest could make a difference *here* but would not make a difference *there* – if all morally relevant factors were the same. Or how could it be that if two candidates have exactly the same attributes *here* then they would be equally suitable for office, whereas if they have exactly the same qualifications *there*, then one candidate would be more suitable for office than the other.

How might the response to this claim proceed?

As a preliminary observation, it does seem that the move for a particular factor '*not making a difference in this particular case*' to '*not making a difference in any case'* seems suspect in day-to-day experience.

Consider raincoats. If the weather is clement, then might be no difference in performance between a cheap raincoat and an expensive raincoat. But can we say this is true of all weathers? No - on a wet day the expensive raincoat performs much better than the cheap raincoat. This may also be true in windy weather or snowy weather. The conclusion from one comparison (in dry weather) cannot be generalised to all weather conditions.

Now consider scientific/medical research. Let's imagine that in a clinical trial it is demonstrated that for a certain population (men under the age of 25) there is no difference in effectiveness or safety between drug A and drug B. Can we say that this is true of all comparisons of drug A and B? Again, no. If the drugs are given to an elderly population, who have different metabolic rates, and different liver function, it may well turn out that Drug A is more effective and has less side effects than Drug B. So again, this might make one reconsider Rachels's generalisation argument - a lack of difference in one situation (young males) cannot be generalised to all situations with the same relevant facts (for example, elderly persons).

So what about moral comparisons such as judgements about the rightness or wrongness of doing and allowing harm? Moral judgement may be different from day-to-day judgement, or from scientific judgement. Rachels thinks this is the case, and there is perhaps a tradition in moral philosophy of universal generalisability. Kopeikin (2020 p177), for example, says '*scientists are aware that their results might not generalise, but most ethicists take their results to generalise universally'*. So conclusions from day-to-day life, or from science, may not be relevant when considering the generalisability of ethical judgements.

So how might one challenge Rachels' moral universibility claim? My approach is to firstly return to Rachels' examples (I've chosen *Frank and John*, but the same sort of argument would work for *Brown and Black*) and to show that whilst Rachels's argument works in some scenarios, alteration of the scenario slightly undermines Rachels's claim. I then offer an explanation (one from Kagan (1988) to provide a rationale for these differences.

Recall the case of John and Frank, where John and Frank are both equally stingy. This observation about stinginess, together with a ceteris paribus claim, allows us to

make claims about the overall moral character of John and Frank i.e. that John and Frank are equally bad overall. Similarly, in the situation where John is stingy, but Frank is even stingier, this will allow us to make the claim that Frank has a worse moral character overall than John. According to Rachels' generalisation principle, we can make a comparison about overall character based on how stingy the individual is, provided that other moral factors are held constant.

Now consider:

1) *Stinginess in a time of plenty*. In this variation, Frank and John are both stingy and as before, Frank is a little stingier than John. But it is a time of plenty, and everyone has everything they could possibly want.
2) *Spoilt children*. In this variation, Frank and John are both stingy and as before, Frank is a little stingier than John. But Frank's children and John's children have become spoilt, perhaps due to the earlier actions of their over-indulgent mother.

In both these two scenarios, the difference in stinginess between Frank and John contributes in different ways to the overall assessment of moral character. In '*Time of Plenty'* stinginess does not matter morally anymore, so the difference does not contribute in any way at all to overall assessment of character. In '*Spoilt children*' the difference between Frank and John is reversed - Frank is now to be praised for being stingier.

Similar arguments can be made for dishonesty. Recall Rachels original example where Frank and John are equally dishonest. In the absence of any other moral differences, Rachels wants to argue that Frank and Johns' overall moral character must be equally bad too. Furthermore, says Rachels, if Frank was a

little more dishonest than John, then this would translate into a difference in moral character (Frank would be overall morally worse than John)

Now, as before, consider two variations:

1) *Mind reading*. In this variation, Frank and John live in time where humans have perfected mind-reading. Everyone knows what everyone else is thinking, so it is impossible to dissemble.
2) *Totalitarian state*. In this variation, Frank and John live in a totalitarian state. Certain members of the community are reviled by the state, and agents of the state seek to find these individuals, performing house to house searches on a daily basis. If these unfortunate individuals are found, then they suffer torture and death.

In the latter two scenarios, the difference in degree of dishonesty between Frank and John contributes in different ways to the overall assessment of moral character. In '*Mind reading* dishonesty does not matter morally anymore, as the whole concept of honesty has become redundant. The difference between the honesty of Frank and John does not now contribute in any way at all to the overall assessment of character. In '*Totalitarian state'* the moral difference between Frank and John is reversed - Frank is now to be praised for being less honest - his dishonesty might allow vulnerable individuals to escape capture.

So even using the examples offered by Rachels himself, it seems that Rachels's generalisability thesis does not go through. The difference between the stinginess of Frank and John, and the difference between the honesty of Frank and John do not contribute in an *invariable* way to an assessment of their overall moral character.

Rachels wants to argue that a difference (or indeed no difference) in honesty or stinginess makes the same difference to moral character *here* as it does *everywhere*. The examples I have offered show that this is clearly not the case; there can be situations ('*Time of Plenty*' and '*Mind-reading'*) where it makes no difference at all, and situations ('*Spoilt children'* and '*Totalitarian state'*) where it makes the opposite difference to the difference suggested in Rachels's examples. Using Rachels's analogies between stinginess/dishonesty and doing harm/allowing harm now allows us to question Rachels about the generalisation of doing and allowing harm. Just because doing the difference between doing harm and allowing harm do not make a difference here, does not allow us to claim that doing and allowing harm fail to make a difference *everywhere*.

2.9. Some theoretical explanations

My argument to date has relied simply on challenging the examples on which Rachels bases his argument. This is reasonable; much of Rachels's argument hinges on the examples he gives. But I now want to offer a theoretical explanation for my response to Rachels.

(i) Kagan (1988) - 'The Additive Fallacy'

Kagan believes Rachels's generalisation argument is faulty. (he calls Rachels's argument the 'Ubiquity thesis', but the claim is identical to what I have called in this essay the 'Generalisation claim'). Kagan thinks that the ubiquity thesis is widely accepted in moral philosophy but is actually false. Furthermore, (suggests Kagan) the ubiquity thesis is false because of the way that reasons in favour or against a moral judgment interact. Rachels (says Kagan) assumes that reasons for and against a moral judgement have to interact in an additive way (simply *adding up* the pros and cons of particular moral options). This incorrect assumption is called by

Kagan the '*Additive Fallacy'*. Kagan concedes that reasons do often interact in an additive way. This seems plausible and is built into our common-sense way of making decisions - we add up the factors in favour of each option and subtract the factors the count against each option. (consider, for example, about the way we choose whether to move to a new job, or take a particular holiday). Whichever option ends up with the strongest positive value is the one we choose. However, says Kagan, reasons do not have to operate in this way. To defend this view, Kagan offers the example of 'Trixie and Fritz'.

Imagine that two individuals, Trixie and Fritz who both have an injury. Let's say that Trixie's injury is considerably more severe than Fritz's injury. Normally (ceteris paribus) the fact that Trixie's injury is the most severe would be a reason for giving priority to Trixie; for treating her before treating Fritz. (or perhaps, in a world where resources were severely limited, for not treating Fritz at all).

Let's now change the context. Now imagine that Trixie is responsible for the suffering of both Trixie and Fritz, and Fritz is not responsible at all (perhaps Trixie was driving too fast and crashed the car, so that both Trixie and Fritz were injured). Responsibility for the injury is an additional moral factor (in addition to the severity of injury) in deciding who gets treated first. If Trixie's responsibility is great enough, then there even may be more reason to treat Fritz than Trixie, even though Fritz's injury is less severe.

There seems to be two ways to reach this conclusion. The additive way (Rachels's way) is to simply add in the contribution from the moral factor of responsibility to the contribution from the moral factor of severity of injury. Possibly this approach works for the case as described above. The moral reasons to aid Fritz rather than Trixie that derive from responsibility, simply overwhelm the moral reasons deriving from the

severity of the injury. But now imagine that Trixie deliberately caused the accident with the *aim* of killing Fritz. She hoped that by crashing the car she could kill Fritz but herself remain unscathed. Unfortunately, the murder attempt went wrong, and Trixie ended up with greater suffering than Fritz. In this situation (says Kagan) you might want to think there is *no reason at all* to treat Trixie. If this is correct (obviously not everyone would agree with Kagan here about the not treating claim, but it does seem somewhat reasonable), then an additive way of moral reckoning starts to look like an inadequate way to make moral judgments. Perhaps the fact that Trixie tried to kill Fritz cannot be represented as an additive reason, subtracting from any reason one might have to treat Trixie on the basis of injury severity. Kagan suggests that a multiplicative method might work better; if the multiplier contribution from the fact that Trixie attempted to murder Fritz was *zero* for example, then there would be no reason at all to treat Trixie. (any number multiplied by zero is also zero). It does not have to be a multiplicative function; more complex functions might be needed for more complex cases. Kagan calls the function that determines how to move from individual contributions to an overall assessment of the rightness or wrongness of a behaviour the *governing function*. And he concludes that '*.... for most factors, their role in determining the overall moral status of an act simply cannot be adequately captured in terms of separate and independent contributions that merely need to be added in*[37],[38].' (1988 p21).

[37] Frances Kamm espouses a similar principle to Kagan, which she calls the 'The Principle of Contextual Interaction'. (Kamm 2007, p17). She does not offer an explanation for her principle however, merely makes a claim about the way that reasons vary in different contexts.
[38] Jonathan Dancy (2004) also makes a claim about the way that moral reasons can alter their strength or valency in different contexts. He calls his approach - value holism. ("a feature that is a reason in one case may be no reason at all, or an opposite reasons , in another' (2004, p7). However he view is more radical than Kagan's who wants to maintain that additivity is not the only way to do moral mathematics.

Kagan's approach provides an explanation for the conclusions I suggested in *Time of plenty/Spoilt childre*n and *Mind-reading/Totalitarian state*. In these scenarios, a simply additive approach would not produce the moral judgement required; more complex manipulation of moral reasons for and against are required.

2.10 Conclusion

I believe I have shown that Rachels's generalisation claim is not successful. Firstly, Rachels fails to explain the outcomes of doing/allowing pair comparisons where there does not seem to any difference between doing and allowing harm. Secondly, his theoretical argument fails to consider the fact that moral reasons can interact with each other and alter their strength. Moral reasons *in favour* of a particular moral judgement may even become moral reasons *against* a particular moral judgement when moved to a different context.

The importance of context will become important as my dissertation proceeds. This chapter is merely a promissory note that having introduced the idea of context, I will expand on the mechanism of how it may influence moral judgements about doing and allowing harm. The challenge posed by Alastair Norcross (2003, p451) retains its importance:

"*It is always possible to object to an argument from a pair of examples that a particular distinction is sometimes morally relevant and sometimes not. The objection by itself, though, has no force. It must be accompanied by an explanation of why the distinction is morally relevant in some cases and not in others*".

The rest of this dissertation will be an attempt to provide an answer to Norcross's challenge.

CHAPTER THREE: DOING AND ALLOWING; CAUSATION AND CAUSAL EXPLANATION

3.1 Introduction

In this chapter, I present a first attempt to explain the moral difference between doing and allowing harm. This account is based on the different way that moral responsibility for harm is explained in cases of doing harm and cases of allowing harm[39]. In short, I will argue that for *doing* harm, moral responsibility is explained (at least in part) by its contribution to causation. Doing harm is blameworthy (at least in in part) because it *causes* harm. By contrast, for *allowings* (which I will argue do not cause anything), moral responsibility for harm is *not* explained by their role in causation (because they do not cause). Allowings (I shall show) do not cause harm, but they do have a role in the causal *explanation* of harm. Both actual causation and causal explanation can ground moral responsibility, but in general, actual causation is a more potent creator of moral responsibility than mere causal explanation. Furthermore, because allowings only causally explain harm, this means that the moral responsibility consequent on an allowi42 of ng of harm is usually less than the moral responsibility consequent on a doing of harm.

So the difference between causation and causal explanation can explain the moral difference between doing and allowing harm. This is not the whole story, however.

[39] As stated earlier, I explicate the moral difference between doing harm and allowing harm in terms of differences in moral responsibility for the harm (an agent has a greater moral responsibility for doing harm than allowing harm). But it could also be explicated in terms of blameworthiness (the agent is more blameworthy for doing harm than allowing harm) or wrongness (it is more wrong to do harm than allow harm) or justification (it is harder to justify the doing of harm than the allowing of harm). I discuss this in detail on Page 15 and 16 of this thesis.

Such an account is unable to explain difference in moral responsibility for different kinds of allowing harm. Later on in this thesis I will introduce the role of teleological explanation in the attribution of moral responsibility for both doing and allowing harm. I shall go on to argue (Chapter 5) that moral responsibility for the doing and allowing of harm can be explained by in terms of a combination of genuine causation and teleological explanation (for doing harm) and causal explanation and teleological explanation (for allowing harm).

Here's how my argument will progress. Firstly I differentiate between causal and moral responsibility. Secondly, I summarise the traditional account of moral responsibility for *doings* – that a necessary[40] condition of moral responsibility for doing harm is that the harm has been *caused* by the doer (no causation, no moral responsibility). Next, I argue that allowings, by contrast, do not cause anything – so causation cannot be a necessary condition for the moral responsibility attributed to *allowing* harms. The claim that allowings do not cause anything is controversial and goes against common sense; we often think of allowings as having the ability to cause harm. (for example, we say that the failure of the gardener to water the plants *caused* them to die).

I argue for the claim that allowings do not cause in three ways - [(a) to (c)].

I claim that (a) the causal relata in causation are events – and as allowings are not events, therefore they cannot cause. Then, using Hall's division of causation (Hall 2004, p252) into 'production' causation and 'dependence' causation, I show that *both* sorts of causation are problematic for allowings of harm. So (b) omissions cannot *produce* anything as they are not involved in transfer of conserved properties

[40] Though not sufficient

(energy or momentum). In addition, (c), there are major difficulties with seeing omission causation as a *dependence* relation. The most pressing of these difficulties is the problem of the profligacy of omissions. If omissions cause by a counterfactual mechanism, then there does not seem to be any principled way to exclude omissions that we do not want to call causal - but which do seem to have the right counterfactual relationship to a particular harm. I discuss several ways that have been suggested to resolve this problem and conclude that none of them are effective.

If omissions do not cause, how do we explain the common-sense view that omissions do appear to be causal? I argue, following Helen Beebee 2004, p291), that common sense is mistaken. In short, when we say that the failure to water the plants *caused* the plants to die, we mis-speak – what we actually mean is that the failure to water is part of the *causal explanation* of the death of the plants. Allowings do not cause anything, but they do figure in the causal explanation of many outcomes.

This allows us to differentiate between moral responsibility for doing harm and allowing harm. The traditional story about moral responsibility for doing harm is correct; *moral* responsibility is grounded in *causal* responsibility – doings are genuine causes of harms. But for the allowing of harm this is not the case. Allowings do not cause harm; at the very most they provide a causal explanation of that harm. And (as we shall see later), an agent is more blameworthy for causing harm than causally explaining the same harm. It therefore seems that there is a difference between the ascription of moral responsibility for doings and allowings, grounded in the relative blameworthiness of causation and causal explanation.

3.2 Causation and causal responsibility

It seems true to say that moral responsibility and causation are related to each other *in some way*[41]. It is therefore important, when thinking about moral responsibility to be tolerably clear about what causation is.

Unfortunately, whilst there are many accounts of causation, there are no clear reductive accounts which are not subject to theoretical problems or counterexamples. Causation can be seen as regularity/constant conjunction (Hume 1888, Mackie 1974), counterfactual dependence (Lewis 1973), physical transference/causal process (Dowe 2000), or probability raising (Hitchcock 2021)[42].

One possible way forward, past this impasse, is to consider a pluralistic view. These too come in various forms, but one interesting and helpful approach is the view, expressed by Ned Hall (2004) that there are just *two* distinct concepts of causation. Hall called these two views, *dependence* causation and *production* causation. Production causation is characterised by theories of causation where there is a spatio-temporal continuous process leading from cause to effect. Dependence causation is characterised by theories of causation where there is counterfactual dependence between cause and effect (such that had the cause not happened, neither would the effect[43]).

Usually, the two forms of causation go hand in hand; for example, if Suzie pushes Tom off the cliff, there is a process leading from Suzie to Tom to the fall, but there is also counterfactual dependence (had Suzie not pushed, then Tom would not fall).

---

[41] At least for doings – I shall say more about allowings later

[42] Even more theories of causation are available – I shall not mention these in this thesis

[43] Illari (2010, p2) provides a nice summary of the way that different theories of causation map on to these two categories. She suggests: **Difference making** – regularity, probabilistic, counterfactual, invariance theories, causal modelling. **Production** – Reichenbach-Salmon mark transmission, Salmon-Dowe conserved quantities theory, Anscombean pluralistic primitivism, Glennan's mechanistic theories.

But sometimes the two can come apart. For an example of dependence causation without production, consider the following examples:

Drop shot

> *Craig teaches Caroline how to execute a perfect tennis drop shot. After the lesson, Caroline can reliably execute drops shots against most opponents.*

Craig causes Caroline to be able to execute a drop shot because this ability depends counterfactually on John's teaching ability – had John not taught Caroline, or taught her wrongly, then Caroline would not be able to pull off a good drop shot. Arguably though, Craig does not *produce* Caroline's drop shot skill – there is no transfer of energy or momentum between Craig and Caroline. It is therefore not quite right to say that Craig *produces* learning in Caroline. But it does seem reasonable to claim that Craig *causes* Caroline's new ability.

Illari (2010, p2) provides another example in the same vein:

Bombing mission

> *Billy and Suzy are pilots in a bombing raid on an enemy city. Suzy flies the bomber to the city and drops the bombs. But Billy is her escort fighter, successfully shooting down all the enemy fighters that threaten to destroy Suzy's plane.*

In this scenario, there is no transfer of energy or momentum between Billy and the target. Yet we want to say that Billy *causes* the destruction of the enemy targets, despite the lack of 'production'.

Conversely, for an example of production causation without dependence, consider the following case:

Firing squad

John is shot by a firing squad. Each member of the 6-man squad shoots accurately and each member's bullet pierces John's heart

In this case, there is no counterfactual dependence between individual members of the firing squad and the death of John. Had any one member of the firing squad failed to shoot (or shot inaccurately), then the death of John would still occur – the other five shooters ensure that this is the case. But it does not seem unreasonable to claim that any individual shooter causes the death of John[44].

Note that the deficiencies in one type of account of causation (production or dependence) are remedied by the other type of account. In that sense the two accounts are complementary. We say that Craig causes Caroline to learn how to perform drop shots because had Craig not taught her, Caroline would not be able to perform the shot. We say that Billy causes the mission to be successful because had he not been there, the buildings would not be destroyed. These are both (successful) dependence accounts. Conversely, we say that the shooters cause John to die, even though there is no dependence on the action of any shooter. Hence the account of causation in firing squad is a (successful) production account. Bearing in mind the difficulties in interpreting causation but working approximately with Halls' two-fold account of causation, we can now say that causation occurs:

[44] Not all philosophers agree that the death by the firing squad is over-determined in the way described in this example. Michael Moore (Moore (2009) p 412, for example argues that events can be what he calls modally fragile. That is, if the events are brought about in slightly different ways (at a different time, in a different position, for example) then they count as different events. So the killing of John by Shooter 1 is a different event to the shooting of John by Shooter 2, which is a different event to the shooting of John by Shooter 3....and so on. If this is correct, then there is no such thing as overdetermination, as each agent does make a difference – if they had not been present then a different outcome would have been produced. It is worth stating that this approach is not universally accepted. Seeing events as modally fragile can produce counter-intuitive results and suggest that (for example) rain that postpones a forest fire from January to February (so that the fire in January is a different event to the fire in February) might (counter-intuitively) be seen as the cause of the fire.

EITHER if a cause *produces* an effect, OR if there is a relationship of *counterfactual dependence* present between the cause and the effect (or if *both* production and counterfactual dependence occurs).

This is a simplification, but a useful one, and one that I will utilise later in this chapter. Note that we are not (yet) talking about *moral* responsibility. Causal responsibility simply identifies the cause of the outcome. Importantly, causal responsibility can be ascribed to moral agents, non-moral agents, or even to physical events. Consider the breaking of a vase. This can be caused by a moral agent (*the vindictive houseguest caused the vase to break*) or a non-moral agent (*the scared cat caused the vase to break*) or even a physical event (*the earthquake caused the vase to break*. The causal responsibility for the vase being broken can be ascribed just as easily to the house guest, the cat, or the earthquake – but we would probably only ascribe moral responsibility for the breakage to the houseguest.

3.3. Moral responsibility for harm

The rest of this chapter is about moral responsibility, so it is necessary to have some idea of the taxonomy of moral responsibility, together with an understanding of the meaning of moral responsibility.

3.4. Prospective v Retrospective Moral Responsibility

Firstly, we need to distinguish between retrospective and prospective moral responsibility. Retrospective moral responsibility concerns what has already happened. We might say, for example, that '*Suzie is morally responsible for the accident*' (because she was driving too fast, say). Prospective moral responsibility

by contrast is forward looking. It is roughly synonymous with 'obligation'. For example, one might say, '*Jack the lifeguard is morally responsible for the safety of the swimmers*'. In this chapter, indeed in this whole thesis, the type of moral responsibility that I will be discussing is retrospective responsibility – responsibility for things that have already happened.

3.5. Direct v derived moral responsibility

A second clarification is necessary. We can be responsible for our acts and omissions, and we can be also responsible for the things that occur in the world because of our acts and omissions. (these are sometimes called outcomes or upshots). The general idea is that outcomes and upshots are produced in some way from our acts and (perhaps) omissions. Responsibility for outcomes is also known as *derived* or *indirect* responsibility, as opposed to responsibility for the act/omission itself, which is also known as *basic* or *direct* responsibility.

By way of example, consider a scenario where Suzie is driving too fast, and knocks over and injures John, a pedestrian. In this scenario:

> Suzie is driving too fast (Suzie's *basic* or *direct* responsibility for driving too fast)
>
> Suzie knocks over John and breaks his arm (Suzie's *derived* or *indirect* responsibility for harming John)

This chapter is going to be concerned exclusively with responsibility for outcomes – or derived responsibility - this is because my topic is harm, and harm is an outcome.

3.6. Reactive attitudes

The most widely accepted account of retrospective moral responsibility for outcomes emphasises that it is appropriate to attribute blame (for wrong behaviours) or praise (for good behaviours).

This way of understanding retrospective moral responsibility derives from Peter Strawson (Strawson 1963/2003) in his article '*Freedom and resentment'*. According to Strawson, if an agent is morally responsible for a wrongdoing, then this provides others with justification for responding to that wrongdoing. The responses to the wrongdoing are emotional reactions which Strawson calls *reactive attitudes* - attitudes like blame, anger, resentment and indignation.

Reactive attitudes are accompanied by the thought that the subject of the reactive attitudes in some ways *deserves* the blame (or anger or resentment). The wrongdoer deserves to be blamed or even punished[45].

Strawson also argued that there exist situations where the reactive attitudes might be suspended and would not be suitable to ground moral responsibility. These might include excusing conditions ('*he didn't mean to do it', or 'there was a much greater harm that could be prevented by doing it'*) or exempting conditions ('*he wasn't himself', 'he was under great stress', 'he's only a kid', 'he has a mental illness'*).

So roughly, according to Strawson, an agent is morally responsible for a harm or wrongdoing if the agent's actions do produce reactive attitudes of blame etc in the agent's community – or would do so if they were aware of the agent's action - and there was no excusing or exempting conditions that meant that the reactive attitudes were not fitting[46]

[45] This works both ways – there will be deserved praise and reward for good behaviour.

[46] The point about excusing conditions needs some clarification. Imagine Harry were to push Joan into a puddle in order to get her out of the way of an oncoming vehicle. Is Harry blameworthy for Joan's ruined shoes? This depends on the seriousness of the harm that has been prevented by the puddle incident – if Joan's life has been saved then we would not consider Harry to be blameworthy at all – indeed we might consider Harry to be praiseworthy. But if the prevented harm is trivial or non-

So, to be clear from now on, when talking about moral responsibility, we are thinking about *retrospective* moral responsibility for harm, that is *derived* (rather than direct) and which is best characterised as producing *fitting reactive attitudes* (blame, anger, resentment etc) in others.

3.7 The Standard Claim Regarding Causal Responsibility and Moral Responsibility – some general comments

The standard understanding of moral responsibility is that causal responsibility is a necessary condition for moral responsibility. So, according to the standard position, *there is no moral responsibility without casual responsibility* (later, I shall argue that this is only true for doings – whereas for *allowing* of harm, causal responsibility is not necessary for moral responsibility – in fact for allowings, I shall argue, there is no causal responsibility at all).

The thought that causal responsibility is necessary for moral responsibility is intuitively attractive. The window is broken. Who is morally responsible? Answer: the individual who broke the window, who caused it to be broken. If it turns out that this individual did not in fact cause the window to be broken (another person broke the window, the wind broke the window...) then the individual would clearly *not* be morally responsible for the breakage. It would be very odd indeed to insist that the individual was morally responsible for the breakage if it could be shown that he was in a different continent at the time of the breakage, had never seen the window, and had nothing at all to do with the breakage and had contributed nothing at all causally to the shattering of the window.

---

existent (perhaps Harry's push stopped Joan from stepping on to a grey paving stone rather than a white one, and this mattered to Harry) then the fitting reactive attitude might still be thought of as one of blame.

Julia Driver expresses this connection in what she calls the *entailment* condition (Driver 2008, p423)

> **Entailment Claim (EC)**: If A is morally responsible for an event[47] e, then A performed an action or omissions that caused e.

Caroline Sartorio (2007) makes a similar claim, but as well as the claim that moral responsibility entails causation, she also makes a claim about explanation – in other words causation explains moral responsibility.

> **Entailment dependence (ED)**: An agent A is morally responsible for an outcome O only if A is causally responsible for O, i.e. only if one of A's actions or omissions caused O; moreover, the fact that one of A's actions or omissions caused O (partly) explains the fact that A is morally responsible for O.

To be clear, these principles imply that unless an agent *causes* a harm, they cannot be held morally responsible for that harm. In other words, the moral responsibility for a harm requires causing that harm.

It is worth noting that Sartorio and Driver assume that both acts and omissions (doings and allowings) cause, and hence both acts and omissions (doings and allowings) are capable of grounding moral responsibility via the mechanism of causation. I will later argue that this is not the case, at least for omissions.

In passing it is important to make clear that causation, whilst *necessary* for at least some forms of moral responsibility (in particular, moral responsibility for doings), is not *sufficient* for moral responsibility.

[47] Driver assumes that moral responsibility concerns events. This is contested (see later in this chapter). It might be better at this point if we simplify as Sartorio does in the next condition and talks about outcomes.

For causal responsibility to be sufficient for moral responsibility, other conditions need to be met. I will not discuss these at length (but I have summarised these in the footnote below) as my focus is on causation as a *necessary* condition for moral responsibility[48]. And to be clear, the argument I will be considering is whether or not it is a necessary condition that an agent be causally responsible for a harmful outcome, in order for the agent to be considered morally responsible for that outcome.

In this chapter, however, I am going to argue against the standard position. More precisely, I will argue that for doings, there is a necessary connection between causing harm and being responsible for harm. But (or so I shall argue) for omissions, this is not true. For omissions, it is not necessary to cause a harm in order to be responsible for that harm. In fact, as I believe omissions do not cause anything, then for omissions, causal responsibility cannot be necessary for moral responsibility.

---

[48] In brief, other conditions for moral responsibility might include an **epistemic** condition, such that for an agent to be responsible for anything, the agent needs to be able to foresee that the harmful outcome would arise. Foreseeing a harmful consequence requires at least that the agent should have the necessary cognitive ability to be aware of consequences, and also has sufficient information in order to be aware of consequences.
If I could not reasonably foresee that singing a song as I walked along would cause the roof girders to vibrate at just the right frequency that the roof would collapse – then I could not be held responsible for the damage caused. Contrast this with driving when tired. I would be responsible for an accident if the tiredness caused the accident, and I could reasonably foresee that the accident would follow from the act of driving whilst tired.
Another condition might be that the harm is brought about **in the right way**. For example, imagine that an evil husband gives his wife a drink of water, imagining (wrongly, as it happens) that it contains poison, and intending that she should die. In fact the wife chokes as she drinks and does in fact die. So an act of the husband for which he is responsible (giving her a drink) has caused harm (death) and this harm was intended by the husband. But one might be reluctant to say that he is responsible for her death – because he has not brought about the harm in the right way (he has brought about her death by a (so-called) deviant causal chain)
In addition to an epistemic requirement, there might also need to be a **control** condition – such that an agent would only be morally responsible for an act or omission that was voluntary or was something that the agent was free to do. If an agent has been compelled to perform a harmful act (for example, due to hypnosis, or by being forced to take a drug that compels me to perform some harmful action, then the agent is clearly not responsible for the harm.

Another way to put this (just to flag this up as the direction in which I hope to proceed) is that the mode of explanation for the doing of harm is not the same as the mode of explanation for the allowing of harm.

I shall argue for this position in some detail later in the chapter. But firstly, for completeness, I will expand and develop the position regarding doings.

3.8 Causal and moral responsibility for doings

For *doing harm*, the relationship between causal and moral responsibility seems to be clear cut. Moral responsibility for *doing harm* seems to *require* causal responsibility. (Clearly, as described earlier, other factors impact on moral responsibility; whether the action was voluntary or not, whether the agent understood the likely outcome of their doing etc. Such factors indicate why causal responsibility is not sufficient (though it is necessary) for moral responsibility). Furthermore, for doing harm, the causal responsibility *explains*[49] the moral responsibility. If Suzie intentionally pushes Tom off the cliff, (a doing of harm) her moral responsibility for the harm to Tom requires that she *causes* Tom's fall; the fact that she caused Tom's fall also explains why she is morally responsible for Tom's injuries - as opposed to Janet, who was walking on the same stretch of coastal path but was not involved in any way in Tom's fall. If Suzie had attempted to push Tom off the cliff, but was an inept pusher, and did not cause Tom to fall (perhaps she completely missed Tom when she tried to push), but Tom later tripped over an exposed tree root and this was the cause of his fall), then Suzie is *not* morally responsible for harming Tom[50]. The causation process does not have to be in its paradigmatic form; for example if Suzie

[49] At least partially, as stated above there may be other necessary conditions
[50] Though she may have been morally responsible for the act of *trying* to push Tom

screamed in a way that she knew would distract Tom, then this would still count as causation.[51]

To summarise the above, it seems reasonably uncontroversial that for doings, causal responsibility is a necessary condition for moral responsibility (though not a sufficient condition, for the reasons (voluntariness, epistemic reasons) given above.

Two final caveats.

Firstly - providing an account of causation of harm, provides (of course) a causal explanation of that harm. We might call this a causal explanation based on genuine causation. This *causal* explanation helps us to understand why there is *moral* responsibility for the harm caused. Later, I shall be arguing that despite not genuinely causing, allowings can produce a (type of) causal explanation for the harm.

Secondly – some doings (and allowings) don't actually cause, but are a constitutive part of the state of affairs that results from the doing. Woollard (2015, p27) gives the following example:

---

[51] Just to lay out my position on production v counterfactual accounts of causation – I believe that doings usually cause by production – though I am agnostic about whether doings *necessarily* cause by production. In the distraction example given above, it could be argued that the scream that distracted Tom caused the fall by a productive causal sequence involving transfer of sound energy from Suzie's vocal cords to the air between Suzie and Tom, followed by the transfer of sound energy to Tom's cochlea and then the transfer of electrical energy from Tom's cochlea to Tom's central nervous system. So again this comes out as a production type of causation. Alternatively, it might be necessary to expand the concept of production to include the transfer of *information* rather than just the transfer of energy/momentum etc (see, for example, Illari and Russo (2014) p 135), Illari, P (2010, p 9). Invoking dependence causation (as in the bombing case cited earlier in this chapter) risks criticism of dependence causation by means of counterexamples involving pre-emption or profligacy. It's worth commenting that although I have written at length about why allowings cannot cause in a dependency way, there are also major problems for dependency causation for *doings* too. This is mainly because of intractable problems with pre-emption (and less because of issues profligacy). If causation is, always a matter of some kind of production (as, for example, Illari (2010) suggests) my account of moral responsibility for doings is similar to the theory offered by Barry, Lindauer and Overland (2014 and Barry and Overland (2016)) where doings of harm involve a complete causal process where a 'complete causal process' involves 'a physical process involving the transfer of energy and momentum from agent to victim' (Barry and Overland, p116). Allowings, say Barry and Overland, lack a complete causal process (p118). My account of doing and allowing includes (as we see in Chapter 5) a central role of intention, which Barry and Overland do not mention.

> Consider a situation in which an agent has promised not to speak. If the agent speaks, the agent's behaviour does not cause the promise to be broken; the connection between the agent's speaking and the breaking of the promise is not a causal one, but something more intimate. As soon as the agent speaks, the promise is broken.

So this provides an exception to my claim that doings of harm necessarily cause that harm. In these sorts of situation there isn't causation; the harm is *constituted* by the doing. So one might modify my claim therefore to say that doings of harm necessarily cause or *constitute* the harm.

3.9 Causal responsibility for allowing – allowings do not cause

The preceding section summarises the story for doings of harm. When we consider the *allowing* of harm, however, the relationship with moral responsibility is different. I want to argue that moral responsibility for *allowings* is *not* grounded in causation. In fact, I want to argue for the stronger thesis that allowings *cannot* ground moral responsibility via causation, because allowings do not cause. I am aware that this claim goes against the prevailing orthodoxy.

I want to present three related arguments for this claim. All three arguments are not unproblematic, and I shall both highlight and respond to these problems as I discuss the arguments. In this endeavour, I might not be completely successful; readers who believe that omissions *do* cause may not be persuaded by each individual argument taken in turn. Taken together, however, the three arguments are mutually self-supporting, and (as I shall attempt to show) provide a reasonably coherent case that omissions are not causal.

Here are the three arguments.

3.9.1 Argument 1 – causation necessarily involves events – and omissions are not events

Firstly, there is a widely accepted theory of causation such that causation is a relationship between events. (Davidson 1967). What are events? Events are positive occurrences with definite spatio-temporal locations. Events occur, have crisp temporal boundaries, (and vague spatial boundaries,) take up time, and may persist as a series of stages. (Lewis 1973).

Here is an example to flesh out the rather abstract account above. Suzie pushes Tom who falls over. Suzie causes Tom to fall. This is a relationship between a (causal) event (Suzie pushing Tom) and an outcome event (Tom falling over). The pushing events takes place at a particular time, as does the falling event. Both events take up time, in this case less than a second. The pushing takes place on the cliff path, with spatial boundaries which are not perfectly defined (Suzie's position changes throughout the event, as does Tom's). There may be a series of stages – Suzie pulling her arm back, Suzie straightening her arm, Suzie pushing, Tom falling, Tom trying to regain balance, Tom eventually falling. And so on.

Events seem the right sort of entities to be part of causation. They are in the world, and exist or obtain at particular times and locations. Helen Beebee (2004, p291) puts this well:

> *I want to uphold the view that causation is a relation between events...I subscribe to what Helen Steward[52] has recently dubbed the network model of causation. The complete causal history of the universe can be represented by a sort of vast and mind-bogglingly complex 'neuron diagram' of the kind*

[52] Steward, H (1997) Chapter 7

commonly found in discussion of David Lewis, where the nodes represent events and the arrows between them represent causal relations....the network model lies behind most of the theories of causation currently on offer; it also lies behind a vast amount of the literature on the philosophy of mind"

Now if this is correct, and causation really is a relation between events, then it seems very problematic to claim that absences can be involved in causation. For allowings, there does not seem to be any event present (and therefore there can be no relationship between events). Consider:

The gardener failed to water the plants.

The lifeguard failed to save the struggling swimmer

Suzie did not push Tom over the cliff.

In these examples, there is just an absence of an event. It is hard to see how there could be any causal relationship when the causal relata are absent.

Sartorio (2007, p752) expresses this well:

According to [some] views, the only things that can be causes and effects are events (positive occurrences with definite spatio-temporal locations) whereas omissions are not events but absences thereof....

The above argument, presented as a deductive argument is:

P1: Only events can cause

P2: Allowings are not events

C: Therefore allowings do not cause

Both P1 and P2, however, can be challenged.

Many philosophers do support P1 - the claim that only events can cause, but not all take this view. Other suggestions for the relata of causation are facts, propositions, tropes, and states of affairs. Perhaps the most popular alternative to events being

part of the causal are *facts* (for example, Mellor 2002). But facts do not seem to be the right sort of entities to take part in causation. Facts do not exist or obtain at specific times or locations. The fact that Suzy pushed Tom, and Tom fell over, is not located at a specific place or time. The fact of Suzy pushing is not earlier than the fact of Tom falling, and the fact of Suzy pushing is not contiguous with the fact of Tom losing his balance. It is not unreasonable to ask, '*How can a fact or a statement push a man ove*r?"

P2 can also be challenged. Some philosophers think that omissions are some kind of 'negative' event. Mellor (2002) takes this line, but it is difficult to imagine what negative events might consist of, or understand how negative events might be causally efficacious.

Another challenge is to suggest that omissions are in some way identifiable with the positive events which take the place of the omission - so the negative event of (say) '*not rescuing the drowning swimme*r' would reduce to, or be identifiable with, the positive event of (say) '*going to buy an ice-cream'*. This is problematic too – mainly because there is no 'unique reduction basis' for the reduction – the omission is multiply realisable with regard to activity, time and spatial location. (Bernstein, (2014) p4). The lifeguard could have been buying an ice-cream, but he could also have been reading a book, chatting to a friend, or playing volleyball. So the negative event of not rescuing the swimmer could reduce to any one of a range of positive events. Nor can we claim that whatever fills the time when the rescue could have been made will suffice as the positive event that reduces to the omission. Imagine that saving the swimmer takes a minute and can be done effectively at any time between 2:00pm and 2:30pm. There is no single one-minute time period to which the swimmer-rescuing reduces, as any one-minute period would have sufficed.

Finally, one cannot identify the place where the rescue could have happened, and call that the omission (something along the lines of 'the positive event that happened at *this* spot'). Clearly the rescue could have happened from the beach, from the sea, from the air etc

Alternatively, one might try a disjunctive version of the positive event strategy – so the negative event of not rescuing the swimmer would reduce to the positive disjunctive event of '*going to buy an ice cream*' OR '*reading a book'* OR '*chatting to a friend*' OR '*playing volleyball'*. But this is problematic too – because this would suggest that causes are disjunctive (with many many disjuncts), and this seems unlikely.

The final (perhaps fatal) problem with seeing omissions as the positive event that was done instead of the omission is that it gets the causal claim wrong. If we want to say that '*the failure to rescue the swimmer caused the swimmer's death'*, then we can't replace this with (for example) '*going to buy an ice-cream caused the swimmer's death*'. It didn't – or at least not necessarily. It is not necessarily true that if the lifeguard had *not* gone for an ice-cream, then the swimmer would have been saved (perhaps the lifeguard could have dropped his ice-cream and rushed across the beach to save the swimmer's death). It is the failure to rescue that caused the death of the swimmer, not the ice-cream purchasing (or whatever other activity that one might wish to put in its place)[53].

To sum up this part of the argument – for the reasons given above, I think it is plausible that omissions are not events, and because of this omissions cannot

[53] In addition, though what have not reached this point in the argument yet, it would be difficult to see how any of these 'substitutions' for an omission would *explain* the moral responsibility for the omission

engage in a causal relationship. Philosophers who deny that only events are involved in causal relations, or believe that omissions are a kind of event, will find this argument less persuasive. I am not relying on this particular argument, though it is congruent with the way I see omissions. I believe that the next two arguments (Arguments 2 and 3 below), particularly in combination, make a much stronger case for the claim that omissions do not cause.

3.9.2 Argument 2 – Causation requires 'push' (or 'oomph' or 'biff')

Many philosophers believe that causation requires some kind of positive driving force – 'oomph' or 'push' or production. And it is difficult to see how omissions can provide this kind of 'oomph'. For example, Armstrong (2014, p175)

> "*Omissions and so forth are not part of the real driving force in nature. Every causal situation develops as it does as a result of the presence of positive factors alone*"

Other philosophers see the idea of 'push' in terms of transmission of energy (or momentum) in the causal process. And, again, it is hard to see how omissions can be involved in the transmission of energy – they are just the wrong sort of entities. A response to the argument from 'push' comes from Ned Hall's, '*Two sorts of causation'* argument. (Hall, 2004). Recall that Hall argues that there were two ways of understanding causation – production causation (described above), and dependence causation (roughly, 'but for' the cause, the outcome would not have happened. So even though we might agree that omissions cannot produce any positive driving force or transfer of energy or momentum, omissions might still cause via the alternative mechanism of counterfactual dependence. So the above argument might rule out causation via production (omissions cannot transfer 'push'

or 'oomph' – what is there to transfer?[54]) but omissions could, perhaps, take part in a counterfactual dependence-type of causation. Unfortunately, counterfactual dependence is also problematic for omissions, as I shall discuss in the next section.

3.9.3 Argument 3 – Omission causation via counterfactual dependence generates a profligate ontology

Perhaps the best-known account of counterfactual causation is the one offered by David Lewis (Lewis 1973, p556)

> *....e causally depends on c if and only if, if c were not to occur e would not occur.*

So let's take the example of the gardener's failure to water the plant as a cause of the plant dying. This makes sense in a counterfactual dependence sort of way - if the failure of the gardener to water the plant did not occur, then the plant would not die. Generalising, for most omissions it is plausible to say that had the omission not occurred, then the outcome would not have occurred either.

Before I discuss the particular problems that counterfactual dependence has for causation via *omissions*, it is worth stating that counterfactual accounts of causation for *both* doings and allowings have significant problems. For example, counterfactual accounts of causation have enormous trouble coping with pre-emption, particularly late pre-emption.

Consider the following case (Menzies and Beebee 2020):

> *Billy and Suzy throw rocks at a bottle. Suzy throws first so that her rock arrives first and shatters the glass; Billy's rock sails through the air where*

[54] Recently,

the bottle had stood moments earlier. Without Suzy's throw, Billy's throw would have shattered the bottle. However, Suzy's throw caused the bottle to shatter, while Billy's throw is merely a pre-empted potential cause. This is a case of late pre-emption because the alternative process (Billy's throw) is cut short by the main process (Suzy's throw) running to completion.

According to most accounts of counterfactual causation, Suzy's throw did not cause the bottle to shatter. Even had she not thrown the rock, the bottle would still have shattered. And clearly, it seems implausible to make this claim; if Suzy did anything she caused the bottle to shatter.

There are responses that can be made to this criticism of counterfactual theories of causation[55]. I simply want to highlight that counterfactual causation is problematic for both doings and allowings. And this has led some philosophers to *deny* that causation by counterfactual dependence is true causation – for example Dowe (2001) calls counterfactual dependence '*Quasi-causation',* and Persson (2002) calls this kind of causation, '*Fake causation'*.

But there is a *particular* and significant problem for explaining causation via counterfactual dependence when it comes to omissions. This is the problem of profligate omissions. Consider the statement:

[55] Such as claiming that events are fragile. So the shattering of the bottle by Suzy's rock is a different event than the shattering of the bottle by Billy's rock, even though the bottle shatters in both scenarios. Such a move avoids the problem of late pre-emption, but creates new problems. If events are as fragile as is suggested above, then this generates lots of counter-intuitive conclusions about causation. Example – a forest fire that occurs in August is delayed by unseasonal rains until January. If the forest fire in August is a different event to the forest fire in January then it seems (counter-intuitively) that the rains caused the January fire – on the grounds that had the rains not come, then neither would the January fire

The gardener failed to water the plant, and the plant died.

Did the gardener's omission cause the plant to die? Well, if we accept causation as counterfactual dependence then yes, for the reasons given in the preceding section. Had the gardener not failed to water the plant, then it would not have died. But if the gardener caused the plant to die by his failure to water, then it seems also to be true that the following statements are true:

The gardener's aunt's failure to water the plant caused the plant to die.

The King of England's failure to water the plant caused the plant to die.

The New Zealand Prime Minister's failure to water the plant caused the plant to die.

The Pope's failure to water the plant caused the plant to die.

In other words, any agent[56] who had they watered the plants would have led to the plants not dying - caused the plant to die.

In fact there are a *very* large number of individuals whose failure to water the plant caused the plant to die. And this seems odd and very implausible. Yet it seems to follow naturally from the counterfactual claim above. The only reason we say that the gardener caused the plant to die is that if he had watered it (if he had not failed to water it) then the plant would not have died. But this is true of the King of England, the Pope, and (literally) millions more. All these individuals have the right pattern of counterfactual dependence with regard to not watering and plant death. It seems implausible to accept that so many different omissions caused this one plant to die

[56] It is worth pointing out that this is probably not necessarily an agent – it could be an automatic water sprinkler, for example, or a leaky roof, on an unusually fierce storm

(and it is even harder to accept that all these millions of people are in some way responsible for the death of this plant).

If it turns out that this is an insoluble problem for counterfactual theories of causation, then it seems (in combination with argument (2)) that omissions do not after all cause anything.

3.10 Responses to the profligate ontology problem

How might we respond to the profligate ontology problem? There seem to be three options:

a) Accept that omissions are profligate but deal with this problem pragmatically.
b) Accept that omissions might be profligate, but suggest some principled way to try and restrict the number of omissions that can be causal – in other words, try to make omissions less profligate.
c) Claim that omissions do not cause anything (my preferred option)

I will discuss these three options in turn. I hope to show neither (a) nor (b) are successful, thus supporting my preferred option, which is option (c).

a) Accept that omissions are profligate but deal with this problem pragmatically

This is the solution offered by Lewis (2000, p75). Lewis tackles the profligate ontology issue by biting the bullet and agreeing that all the profligate omissions we have mentioned so far (the King of England not watering, the Pope not watering etc etc., etc.) are genuine causes. This is because they have the right counterfactual relationship to the harm produced. So it's perfectly true that the Pope's failure to water the plants caused them to die – but we just (for pragmatic reasons) don't talk about it. To quote Lewis (p75) – '*we sometimes refuse to say that an absence*

caused something, even when we have the right pattern of dependence'. To do so would be to violate a conversational rule – for example, that one's input to the conversation should be relevant. To quote Lewis (2000, p 75)

> *There are ever so many reasons why it might be inappropriate to say something true. It might be irrelevant to the conversation, it might convey a false hint, it might be already known to all concerned.*

Here is an example to illustrate the sort of thing that Lewis is getting at. Imagine you walk each day in the park with a friend. You talk of many things, but would not normally say, '*The grass is green'*. It is true that the grass is green, but pragmatically, you would not mention this, because it is both obvious and un-noteworthy. If however, there has been a recent drought, followed by rain, it might then be appropriate to say – '*The grass is green'*. Lewis thinks this pragmatic strategy copes well with the profligate omissions problem. In short, it is true that the Pope's failure to water the plants caused the plant to die – indeed (according to Lewis) every individual who failed to water the plants caused them to die. Pragmatically though, no-one would think it appropriate to mention it. Hence the omissions that do actually cause are *just* the ones that we would think relevant in a particular situation. In this scenario, the relevant omissions would be – the gardener, maybe some friends who have helped in the past. All the other omissions are still causes – but they are causes that no-one would ever think to mention.

There is however a flaw in this pragmatic strategy. It is not just the case that the Pope's failure to water the plants is an inappropriate or an unhelpful conversational gambit; it is actually *untrue* that his failure to water the plants caused their death. It didn't. How could it? – he was hundreds of miles away, busy with important ecclesiastical matters, and totally unaware of the plants. Lewis wants to make *all*

omissions genuine (counterfactual) causes and explain away the intuition that most of these omissions (the Pope, the King, for example) don't seem to be causes at all. He does this by claiming that these non-causes are really causes - but causes that we choose not to mention. But this seems wrong – what seems true is that the great majority of omissions are not causes at all, though a few omissions do *seem* to be causes. I actually think that *no* omissions are causes, and I explain the fact that we *think* that some omissions are causes by arguing that these so-called causes are actually only causal explanations. I thus argue that Lewis's 'pragmatic' solution to the profligacy problem does not work.

In addition (and this will become important as this thesis develops), Lewis's strategy has very little explanatory power. Arguing that the only difference between the gardener's behaviour and the Pope's behaviour is a pragmatic one does not explain why we blame the gardener but would never dream of blaming the Pope.

b) There is some principled way to reduce the number of omissions that can be causal

One popular solution to the problem of profligate causation is to try and restrict the number of candidates for causation by omission. We want a principled way of distinguishing between the omissions of the gardener and the omissions of the Pope – one (we would like to think) is a true cause and one isn't.

The idea here is that omissions can cause - but only *some* omissions, not *all* omissions, are causal. And the omissions that can cause have certain properties that distinguish them from other types of (non-causal) omissions. I will expand on this later in this section, but the properties that are often canvassed to distinguish

between omissions that cause and omissions that don't cause are normative properties – things like '*normalcy*', '*proximity*', '*ought*', and '*reasonable to expect*'. Beebee formalises this idea into two jointly necessary conditions for omissions to be causal: (Beebee (2007, p295)

> *Def: The absence of an A-type event caused b if and only if:*
>
> *(i) b counterfactually depends on the absence: Had an A-type event occurred, b would not have occurred AND*
>
> (ii) *The absence of an A-type event is either abnormal or violates some moral, legal, epistemic, or other norm.*

I now want to argue that there are two serious problems with this strategy.

The most serious problem concerns the nature of causation. By adopting this strategy, it appears that we end up trying to define causation by omission (at least in part) in terms of *normative* concepts (normality, proximity, fault etc). But it does seem very odd to want to include normative factors in the explanation of causation. How could such a paradigmatically physical process as causation be explained or understood in terms of normative or even moral notions? After all, causal processes can occur everywhere in the physical world, in ways that are completely unrelated to the establishment of (human) norms and morality.

As Beebee (2004, p296) says:

> *"if we take the definition to give the truth conditions of causation by absence claims, then causal facts about absence depend in part of normative facts.....but nobody within the tradition of metaphysics of causation that I'm concerned with here thinks that causal facts depend on human dependent norms"*

Secondly, there seem to be specific problems with each and every one of the differing normative concepts that have been suggested. No normative concept seems to be able to differentiate successfully between omissions that cause and omissions that don't cause. I will review four commonly employed additional concepts and show that they do not work. Clearly further concepts could be devised and suggested, or combinations of concepts could be introduced. But overall, I think the chances of success in this venture are low.

So here are four moves made in the literature, together with my response to each move.

(i) **Moral 'ought'**. The idea here is that (for example) the gardener *ought* to water the plants (because he is an employee, because he has promised etc etc) whereas the Pope (of course) is not in this situation. Hence the gardener's conduct is the cause of the plants' death, and the Pope's conduct isn't. That seems reasonable and gets the gardener/Pope difference exactly right. But this can't explain *all* causation by omission. So-called causation by omission often involves non-moral causes – consider the failure of the alarm clock to ring as a cause of being late for a meeting, or the failure of the Gulf stream to move as a cause of ice cap melting. And not all causal omissions involve moral norms. Sarah McGrath (2005, p136) offers a nice example, where there is causation by omission by an individual without the involvement of any breaking of moral norms – "*it is not the case that the hit man (morally) ought to have prevented Mr Big from boarding the plane; nevertheless, his not shooting Mr Big was a cause of the boarding*".

(ii) **Reasonable expectation versus Significant/Surprising**. The idea here is that (for example) there is a reasonable expectation that the gardener will water the plants (because he is employed, because he has promised to, etc etc) but for the Pope there is no such reasonable expectation (he lives in Italy, is not particularly interested in gardening etc). So for the gardener *not* to water the plants, or for the Pope to appear and water the plants, would be '*significant, surprising or the like'*. (Bennett 1995 p133). By contrast, it would not be surprising or significant if the Pope did not water the plants. That again, gets this particular case exactly right. But let's imagine a situation where the gardener (perhaps because he wishes to go on holiday) creates a belief, based on false testimony, that the Pope will water the plants (he's visiting a friend in the next street, say). The plants die – who has caused their death? Not the Pope, despite the plant owner having a reasonable (epistemically justified) expectation that His Holiness would drop by with a watering can. If anyone caused the plants to die, it was the gardener, though on this occasion there was no reasonable expectation that he would water the plant (because everyone was expecting the Pope to come, not the gardener)

(iii) **Normality/'supposed to'**. The idea here is that the gardener *normally* waters the plants, whereas the Pope does not normally do so. In the words of McGrath (2005, p136) the gardener is a '*normal would-be preventer'* of the plant's death. Put slightly differently, but conveying the same idea, the gardener is *supposed* to water the plants, whereas the Pope isn't. A related idea is Judith Thomson's idea of '*fault'* (Thomson 2003, p90). It was the gardener's *fault* (because he was supposed to

water) that the plant was not watered; it was not the Pope's fault that watering did not occur. The concept of 'fault' can deal with the alarm clock problem – it is not unreasonable to say that it was the 'fault' of the alarm clock (in Thomson's terms, the alarm clock was 'defective') that the worker overslept (though this argument is perhaps less convincing for the non-movement of the Gulf Stream). This again gets the gardener/Pope difference right, but there are several problems with this suggestion. Firstly, there is a degree of vagueness in the idea of 'normal'. What about a friend who (Beebee, 2004, p297) waters the plants as a favour but invariably stops doing this after a month or so. If she fails to water the plants, is this normal or abnormal? Secondly, (Sartorio 2007, p756) what about causation by *commendable* omissions? Say I fail to share some salacious gossip about a rival for an academic post. My omission (not sharing) causes my rival to get the job. I have caused this, but it wasn't normal to share the information, and I am not a 'normal would-be preventer' of his reputation remaining untarnished (nor was I supposed to share the information). Nor could I be described as being 'at fault' for not sharing this information.

(iv) **Proximity/closest possible world**. One different way of understanding an omission is in terms of closeness of worlds. One way of reducing the number of omissions that count (counterfactually) as causes is to restrict putative causal omissions to those that occur in a world close to actuality. In the actual world the plant died because it was not watered. To find out whose failure to water the plant was causally relevant, we determine *who would have watered the plant in the closest possible world to our own*. It is

easy to see that the closest possible world to our own in which the plant would have been watered is one where the gardener went about his usual routine. Therefore it is the gardener whose omission in causally relevant. By contrast, a world in which the Pope failed to do the watering is a world that is *very* different from the actual world. For example, in this very different world, the Pope would need to know that the plant needed to be watered, would have to be in the right geographical space, would have to have time and motivation to water....and so on. So the closest possible world account gets the difference between the gardener and the Pope just right. But even this account can be problematic. What if the gardener planned (but then forgot) to ask Glenda, the next-door neighbour (who lives very close to the plants) to do the watering. Glenda would have agreed if she had been asked, but she wasn't (the gardener forgot to ask). Therefore, according to the closest possible world argument, the closest possible world in which the death of the plants is prevented is the world in which Glenda is contacted by the gardener, who then reliably does the watering. On this basis, it is Glenda who causes the plant to die. But this seems wildly counter-intuitive – it was the gardener's original omission to tell Glenda, rather than Glenda's neglect that seems at fault. This type of critique highlights the huge difficulty in making 'possible world' arguments when it is difficult to say exactly which is the closest possible world to an actual world. This particular argument, for example, relies on the claim that the closest possible world to a world in which the plants get watered is a world where Glenda does the watering – rather than the world where Glenda does the watering because the gardener remembers to ask her.

> But this further claim is contentious – it would be easy to argue that the converse was true, and it would be difficult to produce a knock-down argument in favour of either claim. This (as well as a generic suspicion about the use of possible world arguments) undermines the validity of using this approach to solve the profligacy problem.

I suggest, therefore, that each of the above strategies fail, and it seems to me likely that whatever additional constraint is added, the problem of profligate ontology will not be solved (new scenarios will be developed to undermine the strategy).

In addition to the failure of these strategies to solve the profligate omission problem, there is an even more important reason why we might not want to pursue these strategies to eliminate profligate omissions. This is because it seems very implausible to argue that causation depends in any way at all on normative factors (like normalcy, moral ought etc). Causation seems to be out there in the world, and can occur in the total absence of moral beings (the earthquake caused the magma to flow, the rainfall caused the river to flood...and so on). So to introduce normative constraints seems out of step with this understanding of causation.

By elimination therefore, we are left with the third option (my preferred option) – that *omissions do not cause*. This seems to fit the ontology of omissions nicely but leaves one major worry – *the fact that people seem to think that omissions do cause.*

3.11 The commonsense view that allowings do cause.

In an earlier section, I presented several arguments in support of the claim that omissions are *not* causal.

To re-cap, these arguments are that a) causation seems to involve events, and omissions are not events, b) causation requires causal 'oomph', and omissions cannot provide 'oomph', c) causation by counterfactual causation would lead to a profligate ontology. Arguments (b) and (c) need to be taken together – omissions can *neither* provide causal 'oomph', nor provide an adequate causal justification – so cannot cause anything.

But there is one obvious and compelling counterargument against this claim that has not yet been considered. This is the common-sense view that omissions and absences *do* in fact cause. We say (as in the example earlier in this chapter) that the gardener's failure to water the plants *caused* their death - in the same way that we would say '*pouring acid on the plant caused their death'*. Yet, '*failing to water'* is an omission – something is not done (the watering) – and the gardener who omits to water allows the plants to die. Similarly, we say that the doctor's failure to treat the patient for a treatable condition, *caused* the patient to die. Or that the absence of a safety-net caused the trapeze-artist to fall to the ground. And so on[57] Nor is this absence-causation talk confined to ordinary folk - most philosophers take it as an agreed fact that absences do cause. For example, Fiona Woollard claims (2022) that causation applies to both doing and allowings, and this, in her view, means that we need to look elsewhere to find causal (and moral) differences between doing and allowing harm:

> *'One natural suggestion is that the agent who does harm causes it to occur; whereas the agent who allows harm doesn't cause it, but simply fails to prevent it where she could have done so. This suggestion has immediate*

[57] Reiss (2019, p25-26) gives a long list of absences which causes things - from the following academic specialities – physics – marine geology – biology – physiology – psychology – economics – sociology – international relations – development studies – world history

> *moral implications. It seems true by definition (almost) that you can be causally responsible only for upshots that you cause. And it is arguably true that you can be morally responsible only for what you are causally responsible for. This argument gets into trouble if we reflect on the fact that we are often responsible for upshots we allow: the death of the houseplants or the child's illiteracy. When we notice that, in these cases, the plants die or the child remains uneducated because of some failure on the agent's part, it becomes clear that the agent does, in some sense, cause the upshots'.*

But if what I have said in previous sections is correct, then there is *no* causation going on in these sorts of cases – despite the expressions quoted above suggesting that it is a commonplace for philosophers and ordinary folk alike to talk about absences being causative. So if omissions really *don't* cause (as I am claiming), then I need to offer some kind of account of why people commonly speak as though absences *do* cause. This is what I shall attempt now.

One promising approach, suggested by Beebee (2004), relies on the distinction between ***causation*** and ***causal explanation***. The basic idea, which I will expand on and defend below, is that absences/omissions do not *cause*, but absences/omissions do *provide a causal explanation* for outcomes. Thus (according to Beebee) it is *false* to say that the failure to water the plants caused the plants to die, but it is *true* to say that the failure to water the plants is part of the causal explanation of the plant's demise (not: '*the failure to water caused the plants to die'*, but '*the plants died because they were not watered'*) This is the first part of Beebee's argument. The second part is to claim that individuals (philosophers and ordinary folk alike) confuse these two concepts. When they *say*, '*the failure to water the plants caused the*

plants to die', they *actually* mean – '*the failure to water the plants is part of the causal explanation of the plants dying*'.

Beebee's claim is an important one, as it offers an interesting explanation as to why we commonly speak of absences causing. But there is a lot of work to do before we can make good on this argument. *Firstly*, we need to know what causal explanations amount to, and how they differ from accounts of causation. *Secondly*, we need to know how anything can be part of the causal explanation of an outcome, without causing it. *Thirdly*, we need to know how understanding the importance of causal explanation might help with the problem of profligate causal omissions that I mentioned earlier in this chapter. How exactly does an understanding of causal explanation advance our understanding of profligate omissions? Why is there a difference between the causal explanation of the death of the plants related to the failure of the gardener to water them - and the causal explanation of the death of the plants related to the failure of countless distant others to water them? (King Charles III, Pope Francis, say). After all they all stand in exactly the same counterfactual relationship to the death of the flowers. *Finally*, we need to see if Beebee's claim, that folk actually confuse causation and causal explanation, is plausible.

These requirements shape the form of this section.

Firstly, I will outline what causal explanation is and how it differs from genuine causation. Secondly, I will develop an account of how causal explanation can be about causation, but not actually cause.

Thirdly, I will show how the difference between causation and causal explanation can give a solution to the problem of profligate causes.

Fourthly, I will assess Beebee's claim about folk confusing causation and causal explanation.

Finally, I will briefly outline the implications of these ideas for the moral difference between doing and allowing harm.

3.12 What is causal explanation and how does it differ from causation?

As I develop this account of causal explanation, I shall try to bring out difference between causation and causal explanation. This is important; Beebee's account relies on demonstrating a difference between the two concepts.

Let's start with causation. Beebee (2004, p301) begins her article with what is generally accepted to be a key part of the standard account of causation, that it is a relationship between *events*:

> *Causation is a relation between events rather than facts – in other words, the canonical form of causal statements is simply "c caused e" where c and e are events and caused is a two place relation, rather than "E because C" where C and E are facts and "because" a sentential connective.*

To expand, causation is a two-place relation between cause and outcome, both of which are events. I share Beebee's view that causation is a relation between events, and have argued for this claim earlier in the chapter.

By contrast, causal explanation is *not* a relationship between events. The concept of causal explanation was first suggested by Davidson (1967) and developed by Lewis

(1986) and Beebee (2004). Davidson agrees that omissions do not cause, but they are able to provide a *causal explanation* of events:

> *Explanations typically relate statements, not events. I suggest therefore that the 'caused' of sample sentences[58] in this paragraph is not the caused of straightforward singular causal statements, but is best expressed by the words causally explains.* (Davidson (1967) p703).

This idea was further developed by David Lewis (1987, p 217), who thought that '*to explain an event is to provide some information about its causal history*'. For Lewis, a causal history is the complete history of the various causes that produce an outcome. For example, a car collides with another vehicle. What is the causal history of the accident? One could mention the icy road, or the bald tyres, or the tired driver, or the oncoming vehicle. And each of these in turn has a causal history – the tired driver has a causal history including late nights at work, an unhappy baby, money worries, and so on. The icy road has a causal history including a recent snowfall, a sudden drop in temperature, and so on.

Lewis therefore sees any particular event that one might want to explain as standing at the end of a long and complex causal history – made up of a series of events.

To provide a causal explanation then, is to provide *explanatory* information about the causal history of the event (in this example, the accident).

One way (in fact, the obvious way) of providing a causal explanation is by providing a straightforward account of causation. '*The driver braking suddenly caused the*

[58] Davidson has in mind (amongst other examples), "*the failure of the sprinkling system caused the fire" and 'the fact that the dam did not hold caused the flood*" (Davidson 1967, p703)

accident' – '*The cat knocking over the vase caused the vase to shatter'* – '*The movement of the tectonic plates caused the earthquake'*. All of these examples provide a causal explanation by offering an account of causation (of the accident, the breaking of the vase, the earthquake). Plausibly, all accounts of causation are also causal explanations.

But that is not the only way of providing a causal explanation. How else might that be done? As Beebee says (2004, p301) – '*not all casual explanations are reports of causation. The explanans of a causal explanation need not stand to the explanandum as cause to effect'.* This is initially puzzling – how can an explanation be a *causal* explanation without being about causation. In the next section, I will say more about this apparent paradox, and flesh out the concept of causal explanation.

3.13 How can causal explanation be an explanation about causation but not necessarily be causal?

DH Mellor poses this same question quite starkly (1995, p130) '*how can facts explain other facts causally without causing them?*' Or, in terms of absences, how can one claim that absences are causally explanatory, but deny that they are causes?

Beebee has a response to that question (2004, p302):

> '*the most important features of Lewis' account of explanation is that it does not amount to the view that every explanation involves picking out a cause (or some causes) of an event; the way in which causal facts enter into an explanation can be more complex than that. One can give information about an event's causal history in all sorts of other ways – by saying, for instance,*

that certain events or kinds of event do not figure in its causal history, or by saying that an event of such and such kind occurred rather than some particular event.....something can be the explanans of a causal explanation without itself being a cause of the event cited in the explanandum'

I need to expand Beebee's suggestion and outline ways in which causal explanations might explain, without necessarily causing. This is important – in order to make a case that there is such a thing as causal explanation, it needs to be given substance. So here are four categories of examples where (I believe) there is causal explanation going on, but not causation.

i. Exclusion of alternative hypotheses

We can provide a causal explanation of an event by *ruling out* alternative hypotheses about causation. Such an explanation does not offer a cause, but simply explains what *doesn't* cause the outcome.

Here are two examples from David Lewis (Lewis (1986) p220)

> ***CIA man***. *"Why was the CIA man there when His excellency dropped dead ? Just a coincidence believe it or not"*

Here the information is negative and shows that a certain pattern of events – namely a CIA organised assassination – *does not* figure in the causal history of the ambassador's death.

Dormitive virtues[59]*. "the doctor explains that the opium has a dormitive virtue.....even this is not altogether devoid of explanatory information....the test is that it suffices to rule out at least some hypotheses about the causal history of the explanandum. It rules out this one: the opium merchants know that opium is an inert substance, yet they wish to market it as a genuine soporific, so they keep close watch and whenever they see a patient take opium they sneak in and administer a genuine soporific".*

Here the information provided is minimal (and is probably tautologous) but is still able to rule out that one explanation of the soporific event (a trick perpetrated by the opium merchants) does not figure in the causal history of the individual's sleep. Both these cases are thus examples of causal explanations that are not about causation – they give explanatory information about the causal history without being about causation. In other words, they meet Mellor's challenge.

ii. Wide disjunction

Coma*: 'The boy is in a coma because of either an infection, a head injury, a poison, or a haemorrhage'.*

This provides a causal explanation (something caused the boy to be in a coma) but it is too widely disjunctive to count as a cause. If causation is about events, then there is no event here – the options for the possible diagnosis are just too disjunctive to count as an event. But there is plenty of information offered as explanation, so the statement is a causal explanation, but not an account of causation.

---

[59] A term first coined by Molière in the play 'The Imaginary Invalid'. In this play, Moliere makes fun of a group of doctors who explain that the sleep inducing properties of opium are due to its 'Virtus dormitiva' or 'dormitive virtues'.

iii. Absence causation

'The plants died because the gardener forgot to water them'

This brings us back to my overarching argument – that absences or allowings do not cause but provide causal explanation. In the example above, the statement provides causal explanation by explaining that the causal history of the death of the plants did not include an event of the gardener watering the plants. In addition, another causal explanation is available - that in a possible nearby world where the gardener *did* water the plants, then the plants might not have died. Both of these are explanations about the causal history of the death of the plants. This is modal information, and is explanatorily relevant to the death of the plants, but is not an account of causation.

iv. Uncaused events

It is not clear whether uncaused events exist, but there is evidence from quantum physics that some quantum events are uncaused. If uncaused events do exist, then such an explanation ("Event x is uncaused") would count as a causal explanation without being an account of causation.

So causal explanations can be truly causal even though they do not mention causation - causal explanations give information about the causal history of events, even though the explanans does not stand in the same relation to the explanandum as cause to effect.

All these four types of explanations thus offer information about the causal history of an event without causing that event. This seems a plausible answer to Mellor's (1995) question - '*how can facts explain other facts causally without causing them?*'

The facts of causal explanation contribute to understanding the causation process without causing it.

Beebee (2004, p 281) sees this specifically in terms of a modal relationship. She gives the example of a gardener ('Flora') failing to water plants, with the outcome that the plants eventually died:

> '*when I say that the orchids' death occurred because Flora failed to water them, you learn something minimal about the death's causal history; that it did not include an event of Flora watering the orchids. But you also learn something about the causal structure of nearby worlds where Flora didn't fail to water the orchids – namely that the causal processes that ensued at those worlds did not cause (or perhaps might not have caused) the orchid's death'.*

This is consistent with (though not exactly the same as) Lewis's earlier account of causal explanation – that causal explanations work by ruling out incorrect hypotheses (and also by citing general causes rather than specific causes – but I won't discuss the latter explanation in this essay).

This may seem a rather poor kind of causal explanation – would it not be better to provide the correct hypothesis, rather than just rule out unsatisfactory hypotheses. But it is still a causal explanation *of sorts*, even if there are better forms of causal explanation available. Indeed, ruling out a possible but serious possibility may be crucially important in some situations. Consider:

> Jane has a febrile illness with cough, loss of taste and smell. At the time of her symptoms, there is a Covid epidemic, and those around her are being admitted with respiratory conditions, some of which turn out to be fatal. So Jane is desperately worried she may be suffering from Covid-19. She takes a Covid test, and the (reliable) test is negative. This provides a (admittedly

negative) causal explanation of Jane's symptoms ('*it's not Covid*)' but in fact it is the only explanation that matters. The exact diagnosis (common cold perhaps) does not matter – what it is *not* matters enormously.

Granted an account of actual causation ('*Jane's symptoms were caused by coryza'*) would be better because this diagnosis not only rules out Covid but gives useful information about what is exactly wrong ('*it's only Coryza, so it's fine for you to come and visit'*). But the crux of the causal information is also provided by ruling out the only diagnosis that really matters to Jane. (Covid).

So, despite the initial oddness, it does seem there is something going for causal explanation as a workable explanation of the relation between allowings and outcomes. The allowing does not *cause* the outcome but is involved in the *causal explanation* of the outcome (with causal explanation understood in the Lewis/Beebee way).

To put this another way, information about the allowing answers a 'Why' question with a '*because'* statement, rather than an answer which involves causation. So, rather than saying that:

(A) The gardener not watering the plants *caused* them to die, we say

(B) The plants died *because* they were not watered by the gardener.

Statement (A) makes a *causal* claim and is false. Statement (B) makes an *explanatory* claim and is true – it provides a explanans for the explanandum that the plants had died.

So (as an initial hypothesis, and to summarise where we have got to) doings cause harm (they are involved in the causation of a harmful outcome), whereas allowings are only involved in the causal *explanation* of a harm. Causing harm is also a way of providing a causal explanation of that harm, so consequently doings both cause and

causally explain harmful outcomes. By contrast, allowings only causally explain (in this very broad 'Beebee/Lewis understanding of causal explanation) – but they do not actually cause. If we accept this, we have a conceptual difference between doing and allowing harm.

3.14 Can causal explanation help us understand profligate omissions?

My claim is this. I have argued that absences do not cause, but they do provide a causal explanation of an outcome. The relationship is not the causal explanation - '*c causes e'* (because absences do not cause) - but the causal explanation – '*E because of C*'.

Can this difference – between cause and causal explanation, differentiate between doing and allowing harm? Beebee (2004, p306) thinks it can. To do this she proposers a *saliency* argument. I actually don't believe that the salience argument works, but it is instructive to reflect on why this argument does not work, and this in turn points us to a better argument.

3.15 Beebee's salience/adequacy argument

Beebee introduces an argument that attempts to differentiate between genuine absence causal explanations and spurious ones. She denies (as do I) that absences can cause so claims that the statement '*The gardener caused the plants to die by failing to water them'* and the statement, '*The Pope caused the plants to die by failing to water them*' are both false. But she thinks the causal explanation – '*The plants died because the gardener failed to water them*' is true – as is '*The plants died*

because the Pope failed to water them'. They are true because they are both explanatory statements rather than causal ones.

There is one obvious problem, however. If they are both true causal explanations, how can one differentiate between the two, with the ultimate aim of comparing moral behaviour for the two conducts (we want to say that the gardener is blameworthy for the death of the plants, but we don't want to say that Pope is similarly blameworthy). Beebee suggests that one way to do this is to by comparing the *adequacy (or salience)* of the two casual explanations. For Beebee, causal explanations need to be both true *and* adequate. A good explanation, Beebee, argues, will allow us to decide who is to blame for the outcome.

This view has a lot going for it. For causation, it seemed illegitimate to use norms to differentiate between two absences which have the same counterfactual relationship to the harm produced. But for causal explanation, which is an epistemic notion, this problem does not apply – it seems very appropriate to talk about salience in this context. One causal explanation can be more salient than another causal explanation.

So if the plants died because the gardener failed to water them, then this is a very salient explanation – it allows one to apportion blame for the death of the plants. By contrast, if the plants died because the Pope failed to water them, then this is a very non-salient explanation. It would be very confusing to ascribe the plant's death to the Pope – he lives miles away and has no connection with these plants. It seems therefore that the gardener's omission is a *much* more salient causal explanation of the death of the plants than the explanation suggested by the Pope's omission. Indeed, if asked for an explanation of the death of the plants, mentioning the Pope would be a ridiculously non-salient explanation to offer. By contrast, the omission of

the gardener (who was employed by the owner of the plants, had promised to water them, and lived close by) provides a remarkably salient causal explanation of the demise of the plants.

Stephen Mumford (2021, p84) seems to agree with Beebee's line of reasoning – he contrasts the acceptability of a salience argument for causal explanation with the acceptability of a causal explanation for causation:

> *'But given that in neither case is it allowed that failure to water the plants killed them – that is caused them to die – then a normative reason why one counterfactual is more relevant, stable or normal than the other is entirely acceptable. This does not commit us to a normative conception of causation itself.'*

In passing, it is worth clarifying that salience in a causal explanation is clearly contextual. For a causal explanation to be salient, it needs to address the particular needs of the agent to whom the causal explanation is being presented.

As Varzi (2007, p156) explains:

> *When reporting causal transactions, we are generally interested in explaining why something happened, and the adequacy of a causal explanation does not only depend on the truth of what is said. When it comes to an explanation, we also must make sure we describe the cause in a way that proves informative given the relevant background of shared knowledge and presuppositions.*

To give an example. In the account above, I have assumed that the causal explanation involving the gardener is the salient one, and the causal explanation involving Pope Francis is the non-salient one. Doubtless, given the differing circumstances and commitments of the gardener and the Pope, this will, in general

be a fair assessment. But this is not necessarily the case. It may be that one particular recipient of a causal explanation is particularly interested in the promise that the Pope made to support UK homeowners with their horticultural needs and responsibilities. Maybe for *that* group of individuals, a causal explanation about the Pope's watering activities would be extremely salient ("*Why didn't the Pope water the plants – he claimed he was interested in them?*") Or perhaps a close friend of the gardener, dismayed about the increasingly onerous plant-watering duties placed on him by her employer, would not be interested in any causal explanations of the death of the plants that involved the gardener ("*Why didn't I water them – because I've too much to do already. It makes more sense for the Pope to do the job*"). Doubtless these circumstances are rather fanciful – but they illustrate that there may be a range of situations where the causal explanation offered is not salient to the particular needs of the interlocutor.

3.16. Problems with the Beebee/Mumford account

Beebee's account is attractive, but I think it does not quite get the link between causal explanation and moral responsibility quite right.

Recall that we are looking for a way to differentiate between the gardener's failure to water the plants and the Pope's failure to water the plants. Beebee claims that the salience of the explanation will do the trick – the gardener's absence is salient, whereas the Pope's absence is not salient (p 307):

> *'...the corresponding explanatory claims.....are both true. However they do not, in most contexts, count as equally adequate explanations...if our explanation is to count as adequate as well as true. [it] must be sensitive to why the explanation was requested in the first place**...**the truth of my utterance*

doesn't depend on the moral question of who is to blame, but the adequacy of my explanation does...'

Hence we say the gardener's failure to water provides a better causal explanation than the Pope's failure to water, and hence we might want to go on and link salience and blame and claim that the gardener's failure to water was more blameworthy than the Pope's failure to water.

But here's a problem with this argument. Yes, the causal explanation involving the gardener is much more salient than the causal explanation involving the Pope. But one might ask - why should that matter for blameworthiness? Why should salience determine blameworthiness in the way suggested by Beebee?

Furthermore, it gets the relationship for between the gardener and the Pope completely wrong. We want to say that the gardener is responsible, but at the same time we want to say that *the Pope is not blameworthy at all*. But the salience story gives us only that the Pope is *less* blameworthy than the gardener, on the grounds that the explanation involving the Pope is less salient that the corresponding explanation involving the gardener. But, according to the counterfactual conditions underpinning the concept of causal explanation, there is still some blame attached to the Pope, because the same counterfactual condition still applies ('*had the Pope watered the plants, they would not have died*).

We don't want to claim that the Pope's failure to water makes him *less* responsible than the gardener. We want to say that the Pope's failure to water does not make him responsible *at all*. So there is something lacking in the causal explanation.

3.17 Confusion between causation and causal explanation

Where have we got to so far? We have introduced the idea of causal explanation, and suggested that although allowings do not genuinely cause, they are involved in the causal explanation of an outcome. This is promising; it suggests a moral difference between doings and allowings based on the difference between causation and causal explanation, though, as mentioned above, there is a difficulty in this account in that it does not explain why there is a moral difference between what the gardener allows and what the Pope allows.

Let's simply note this difficulty for now. There is a further claim made in the Beebee paper (p 305) - that non-philosophers ('the folk') make an error when making statements about causation by absence. Firstly, they *confuse* genuine causation with causal explanation – so what they *call* causation by absence, is actually causal explanation by absence. When they say, '*The failure of the gardener to water the plants caused the plants to die'*, they actually mean – '*The failure of the gardener to water the plants is part of the casual explanation of the death of the plants*'. They move from 'Y because of X' to 'X causes Y' - and fail to notice the difference between the two.

If this confusion is correct, then it could provide a relevant difference between doings and allowings, related to causation but actually about two different sorts of causal explanation. Recall Woollard's complaint (2016) that I quoted earlier in this chapter;

> *One natural suggestion is that the agent who does harm causes it to occur; whereas the agent who allows harm doesn't cause it but simply fails to prevent it where she could have done so. This suggestion has immediate moral implications. It seems true by definition (almost) that you can be causally responsible only for upshots that you cause. And it is arguably true*

> *that you can be morally responsible only for what you are causally responsible for. This argument gets into trouble when we reflect on the fact that we are often responsible for upshots we allow: the death of the houseplants or the child's illiteracy. When we notice that, in these cases, the plants die or the child remains uneducated because of some failure on the agent's part, it becomes clear that the agent does, in some sense, cause the upshots. Moreover, most widely accepted contemporary accounts of causation imply that some event or fact involving these agents causes the deaths or illiteracy.*

So perhaps Woollard is mistaken; what she should have said is that 'it becomes clear that the agent does, in some sense, *causally explain* the upshots'. This makes sense, and, providing there is in fact confusion between genuine causation and causal explanation, can explain the widespread belief that allowing harm can cause harm (as in the gardener/death of plant scenario).

3:18 So what evidence is there that there is in fact confusion here?

Beebee does not in fact try very hard to defend the causation/causal explanation confusion claim. She recognises that it is an empirical claim, as to what people are (2004, p301) '*ordinarily inclined to judge'*. And rather than providing evidence in favour of the claim, she simply asserts that (p301) – *'I have not come across any evidence to suggest that the claim is false'.*

Since Beebee's article was published, there has been (at least) one attempt to investigate Beebee's empirical claim that folk do confuse causation and causal explanation. This was done by Jonathan Livengood and Eduoard Machery, some

three years after Beebee's paper was published (Livengood & Machery (2007), p 107-127). I summarise Livengood and Machery's study design and claims below.

3.18.1 Livengood and Machery's (L&M) 'The Folk Probably Don't Think What You Think They Think' claim

Livengood and Machery (thereafter L&M) attempted to test whether individuals confused causation by absence and absences being part of a casual explanation. More precisely, (L&M p 108) they wanted to see if:

> '*people assert causal sentences - an event c causing the occurrence of event e – in place of explanatory sentences – the occurrence of e being explained by reference to c'.....[this] happens because people confuse causal explanation with causation, and which absence is asserted to be the cause of an event depends on what makes for an adequate explanation of that event. Denial of causation by absence happens because people confuse inadequately explanatory sentences with false ones and because people deny sentences they believe to be false'.*

More specifically, L&M wanted to test if (p108) – '*folk really fail to distinguish causation from causal explanation'*.

They attempted to do this by presenting subjects (actually students[60]) at the University of Pittsburgh) with a series of three vignettes about a young girl climbing a rope. Half of the subjects were asked if they agreed (or disagreed) with a statement that involved causation (the 'causation condition') and half the subject were asked if

[60] The students had an average age of 21, with an age range of 18 - 60, and the gender split was male, 41%, female 59%.

they agreed or disagreed with a statement that involved causal explanation (the 'explanation condition').

Here are the vignettes:

Broken rope case. *Susan had to climb a rope in gym class. Susan was a very good climber. She started climbing, but the rope broke before she reached the rafters. She fell on the ground'*

Causation condition – On a scale of 1 to 7, (1 indicating that you totally disagree and 7 indicating that you totally agree) how much do you agree with the following claim – '*The rope breaking caused Susan to fall*'.

Explanation condition - 'On a scale of 1 to 7, (1 indicating that you totally disagree and 7 indicating that you totally agree) how much do you agree with the following claim – '*Susan fell because the rope broke'.*

Rope case. *Susan had to climb a rope in gym class. Susan was a very good climber, and she climbed all the way to the rafters*.

Causation condition – On a scale of 1 to 7, (1 indicating that you totally disagree and 7 indicating that you totally agree) how much do you agree with the following claim – '*The rope not breaking caused Susan to reach the rafters*".

Explanation condition - 'On a scale of 1 to 7, (1 indicating that you totally disagree and 7 indicating that you totally agree) how much do you agree with the following claim – '*Susan reached the rafters because the rope did not break'.*

Unsafe rope case. '*Susan has to climb an old, worn-out rope in gym class. She wondered if it would support her weight. Susan was a very good climber. Though nervous, she climbed all the way to the rafters'*.

Causation condition – On a scale of 1 to 7, (1 indicating that you totally disagree and 7 indicating that you totally agree) how much do you agree with the following claim – '*The rope not breaking caused Susan to reach the rafters*'.

Explanation condition - 'On a scale of 1 to 7, (1 indicating that you totally disagree and 7 indicating that you totally agree) how much do you agree with the following claim – '*Susan reached the rafters because the rope did not break'.*

Before discussing the interpretation of the results, it is worth exploring some of the assumptions made by L&M in designing this study. The causation condition part of the study seems straightforward – the word 'causation' is in the target sentence. For the explanation condition part of the study the authors believe that using the word '*because'* (as in: '*because* the rope did not break') would capture the idea of causal explanation. I think this assumption is wrong, however, and this seriously undermines the value and the conclusions of the study. The subjects in the study may have assumed that '*because'* also referred to causation. So individuals who agreed with the 'because' statement may have been assenting to a *causation* story rather than a *causal explanation* story.

In addition, even the subjects who correctly linked 'because' with explanation, might have been agreeing to some other means of explanation[61] than a causal explanation, rather than differentiating between causation and causal explanation.

Having said this, the results of this study are interesting.

For the *broken rope case*, the mean score for the subjects in the causation condition group was 5.77. The mean score for the subjects in the explanation condition was 5.47. This difference is significant ($p<0.001$). The score for indifference (neither

[61] See the next chapter of this thesis for other methods of explanation

agree with nor disagree with) was 4.0. So it seems that the subjects in this part of the study

a) were more likely to agree with a causation statement than disagree with a causation statement
b) were more likely to agree with a causal explanation statement than disagree with a causal explanation statement
c) were approximately equally likely to agree with a causation statement than to agree with a causal explanation statement.

This is partially, though not totally, predictable. The *broken rope* vignette (which was introduced as a control) is not about an absence at all – the rope snapping was the cause of her fall, so this is a straightforward causation statement. Nevertheless, it is interesting that even with such an archetypal 'causation' statement, not everyone agreed that the broken rope caused the fall, and an almost equal number of subjects agreed that the broken rope provided a causal explanation of the fall. Again, this throws doubt on the conclusions reached by the authors.

For the *rope case*, the mean score for the subjects in the causation condition group was 2.73. The mean score for the subjects in the explanation condition was 3.5. This difference is significant ($p<0.001$). The score for indifference (neither agree with nor disagree with) was 4.0. So it seems that the subjects in this part of the study

a) were more likely to disagree with a causation statement than agree with a causation statement
b) were more likely to disagree with a causal explanation statement than agree with a causal explanation statement

c) were significantly more likely to disagree with a causation statement than to disagree with a causal explanation statement.

On balance then, the subjects favoured causal explanation rather than causation for the best account of the role of the rope not breaking in the girl achieving success.

For the *unsafe rope case*, the mean score for the subjects in the causation condition group was 3.06. The mean score for the subjects in the explanation condition was 4.0. This difference is significant ($p<0.001$). The score for indifference (neither agree with nor disagree with) was 4.0. So it seems that the subjects in this part of the study

a) were more likely to disagree with a causation statement than agree with a causation statement
b) were equally likely to disagree with a causal explanation statement or agree with a causal explanation statement
c) were significantly more likely to disagree with a causation statement than to disagree with a causal explanation statement.

In this vignette, the absence was made clearly explanatorily relevant. Even then, the subjects were not inclined to assent to a sentence causally explaining an event by reference to this absence (though they were even less inclined to assent to a sentence claiming that the absence caused an event).

Overall then, the L&M experiments do not provide evidence against Beebee's account of confusion between causal explanation and genuine causation. For at least the absences in this study, there was (perhaps contrary to popular belief) little appetite for asserting that absences cause. In favour of Beebee's approach, there was rather more appetite for asserting that the failure of the rope to break provided a causal explanation of the success of the climb. But the difference between

causation and causal explanation is not striking. In addition, there are major problems with the study (discussed earlier) and it may therefore be premature to reach any firm conclusions from the study. What Beebee (and L&M) achieve however, is to throw some doubt on the idea that absences cause – whether there is empirical evidence that absences operate by providing causal explanation is more contentious.

3.19 The moral difference between causation and causal explanation

Let's recap on my claims so far. I have argued that doings cause harm whereas allowings only causally explain harm. Hence doings and allowings are different. Furthermore, folk typically *confuse* causing and causally explaining and this gives rise to the false belief that allowings cause harm. What folk really mean is that allowings *causally explain* harm, which is (as explained earlier in this chapter) a different concept to the concept of genuinely causing harm.

I now want to suggest that causing a harm is morally worse than merely providing a causal explanation of that harm. That an agent who genuinely causes a harm is more blameworthy than an agent who only causally explains a harm. This final piece in the jigsaw will then allow me to argue that doing harm is morally worse, more blameworthy, than merely allowing harm.

Here is my argument for the final part of the claim.

The first strand of my argument is an intuitive claim – it does seem that (as discussed earlier) causation appears to be necessary for the moral responsibility associated with doing harm. To quote Carolina Sartorio (2007, p750),

> *Imagine that someone is killed in a remote part of the world: someone with whom you have never had contact of any sort. Do we consider you*

> *to be morally responsible for his death? Obviously not. Why not? Intuitively, because you weren't causally related to the death in any way: there is no way to link anything you did to the person's death. So here's a very natural idea about the relation between moral responsibility for outcomes and causation: moral responsibility for an outcome requires causing it. Moreover, when an agent is morally responsible for an outcome, the existence of causal responsibility partly explains the existence of moral responsibility: the agent is morally responsible, in part, because he is causally responsible.*

Another intuitive claim - the link between genuine causation and moral responsibility seems much stronger than the link between mere causal explanation and moral responsibility. Just because an agent's action is part of the causal explanation of a harm does not automatically imply (as it does with causation) that that agent is morally responsible for the harm. Similarly - just because the information is explanatorily relevant to the harm does not necessarily mean that a relevant agent (who is involved in the causal explanation) is morally responsible for that harm. This is possibly because the causal explanation for an allowing is only providing information about the causal structure of nearby worlds. As Beebee argues (2004, p305)

> *Well, when I say that the orchid's death occurred because Flora failed to water them, you learn something minimal about the death's causal history; that it did not include an event of Flora's watering of the orchids. But you also learn something about the causal structure of nearby worlds where Flora didn't fail to water the orchids – namely that the causal processed that ensued*

at those worlds did not cause (or perhaps might not have caused) the orchid's death.

One way to look at this is to consider that the causal explanation for allowing is giving the wrong sort of information than is necessary to allow us to ascribe moral responsibility. In a sense it is one step removed from the information necessary to correctly ascribe moral responsibility.

This consideration is linked to the profligacy problem for omissions. Recall that the both the failure of the gardener (who was being paid to water the plants) and the Pope are counterfactually linked to the death of plant. Both the omission of the Pope and the gardener causally explain the death of the plants. But the Pope is clearly not responsible for the plants' death. From this we can conclude that the link between causal explanation and moral responsibility is not strong – just because an agent features in a causal explanation of a harm does not mean that they bear responsibility for the harm that ensues (as in the 'Pope' example). Note that this is not the case for doings – because doings involve genuine causation, there is a much closer link between the doing of harm and the moral responsibility for that harm.

3:20 Which is a more basic distinction – the doing-allowing distinction or the genuine causation- causal explanation distinction?

I have argued that the moral distinction between doing and allowing harm depends (at least in part) on the distinction between genuine causation and causal explanation. Allowing harm is less morally serious than doing harm, at least in part, because allowing involves only a causal explanation, whereas doing involves genuine causation. The causation/causal explanation distinction explains the doing/allowing distinction.

But it could be argued that explicating the doing/allowing distinction in terms of the causation/causal explanation distinction get the dependence the wrong way round. Perhaps the distinction between doing and allowing is the more primitive one, and understanding the difference being genuine causation and causal explanation does not actually help us understand the distinction between doing and allowing harm. Or - perhaps both distinctions are difficult to understand, and one does not depend on the other. Or even - perhaps both distinctions depend on some other (as yet unspecified) other distinction.

I want to argue that the causation/causal explanation distinction is the more basic one. Causation and causal explanation seem to be fundamental metaphysical properties of the universe. It is possible to see how doing (and allowing) depend on causal concepts, but it is not obvious that this relationship travels in the opposite direction. Fiona Woollard (2015) makes a similar point – she denies that the moral difference between doing and allowing is 'unsuitable to be a basic fact' (p5). Nor does she want to say that causation/causal explanation are basic facts – but it appears to me that they are more basic than the doing/allowing distinction. It appears that causal concepts help us to understand the doing/allowing distinction – and not the other way round.

3.21 Conclusion

I think there are two main conclusions from this chapter, but also one problem with the account so far.

First, the positives:

3.21.1 Causal explanation and moral responsibility.

We now have a reasonably satisfactory account of causal explanation. How does this help in understanding the attribution of moral responsibility for allowings? Recall that for doings, genuine causation seems to be necessary (though not sufficient) for moral responsibility. Moral responsibility seems to entail causal responsibility. Here, the linkage between causation and moral responsibility is both intuitive and an easy concept to grasp – if an individual has caused harm, then they are morally responsible (subject to the usual considerations about freedom, capacity, epistemic factors etc). If they haven't caused a particular harm, then, in general, they are not morally responsible for that harm. Such a claim is intuitive and well-established. This connection is actually reflected in many legal systems (see for example, Moore, 2009).

Things are less straightforward for allowings. As I have argued allowings do not cause. In addition, it is uncontroversial that moral agents can be morally responsible for allowings - the gardener is morally responsible for not watering the plants, the doctor is responsible for not treating the patient, the parent is morally responsible for not feeding the child....and so on. How then, do we attribute moral responsibility for allowings? I have argued that allowings of harm can causally explain the harm that follows, even though allowings do not genuinely cause the harm. It seems therefore that there is a difference in kind between doing and allowing, and this difference is related to their respective roles in genuine causation and causal explanation.

Can this difference in kind explain the moral difference between doing and allowing harm, such that do harm (to cause harm) is (at least sometimes) morally worse than merely allowing harm?

Here things are less clear, though one argument might go like this.

3.21.2 Why does it appear that allowings cause?

Secondly, I also claim to have answered the question, (as posed by Fiona Woollard), as to why it *appears* that allowings can cause harm. I argue that this is actually a mistake. Allowings appear to cause harm, I have argued, because they provide a causal explanation for the harm allowed which is then easily confused with actual causation (though, it is no such thing). Causation and causal explanation are easily confused – it is easy to conclude (wrongly, I argue) that allowings cause harm – when in fact they do no such thing (they only provide a casual explanation for the harm allowed).

3.21.3 A problem: differentiating between blameworthy and non-blameworthy allowings

Despite the two positive conclusions cited above, I argue that my account of allowing harm is incomplete – my account fails to explain at least one important thing.
The causal explanation account does not explain differences in attribution of moral responsibility between different types of causal explanation. My account, does not, for example, explain why the *gardener's* failure causally explains the death of the plants but the *Pope's* failure does not . Causal explanation is a counterfactual relationship – and is the same counterfactual relationship for both gardener and Pope. Both the gardener and the Pope failed to water, and the plants died. For both gardener and Pope, the death of the plants could have been prevented if they had watered. So the basis for asserting that the gardener causally explained the death of the plants is identical to the basis for asserting that the Pope causally explained the death of the plants. In this sense, there is no difference between gardener and Pope. And whilst it seems plausible that the failure of the gardener to water causally

explains the death of the plants, the same story applied to the Pope is *far* less convincing (in fact not convincing at all).

This difficulty becomes more pressing when we consider *blameworthiness* for the death of the plants. We very much want to assert that the gardener was blameworthy for the death of the plants, but we certainly do not want to make the equivalent case for the Pope. We actually want to make rather a complex claim – we want the gardener to be blameworthy, but we want the Pope (and other unrelated individuals who also could have watered) the plants to be non-blameworthy. We cannot however separate the two – either both gardener and Pope are blameworthy (because they both have a counterfactual link to the death of the plants) or neither of them are blameworthy (because neither causes, they only causally explain). The destinations we want to get to do not seem (yet) to be available to us.

As Sarah McGrath (2005) says:

> *Another difficulty for the view is that Barry, and not Carlos, is responsible for the death. Causation and responsibility are linked, albeit in some hard to spell out way; intuitively the fact that Barry, and not Carlos, is responsible is linked to the fact that Barry's not watering, and not Carlos's, is a cause. But this link disappears if neither omission is a cause, and the two differ just in that one and not the other is explanatory.*

So in order to develop a complete account of the wrongness of allowing harm (and hence by implication) the difference between the wrongness of doing and allowing harm, we need to do some more work. What we have so far is promising, and explains part of the story; but it is still not complete.

I will present the full story in Chapter 5. Bur first (Chapter 4) I want to make a case for other forms of explanation than causal explanation before going on to argue for

the importance of intention in explaining the moral difference between doing and allowing harm.

3.22 A note on terminology

In this chapter I made a distinction between doings which cause, and allowings - which merely causally explain. Of course doings causally explain too – they do this by causing (unlike allowings, which *only* causally explain). To avoid confusion, from now on, when I speak of doings, I will use the term *genuine* causation, and then I speak of allowings I will use the term causal explanation. So I will say:

Doings genuinely cause

Allowings causally explain

CHAPTER FOUR – OTHER TYPES OF EXPLANATION

4.1 Introduction

In the previous section, I argued that the so-called *entailment claim* about moral responsibility and causation is not completely accurate.

The entailment claim states that:

> *'If A is morally responsible for an event e, then A performed an action or omission that caused e.* (Driver 2008 p 423)

I suggested that this was true for actions, but not true for omissions. I argued that the following claim is more precise:

> *"If A is morally responsible for an event e, then A performed an action or omission that was part of the causal explanation of e".*

I then wondered then if *explanation* is doing the work here – that it is *explanation* (causal or not) which is necessary for claims about moral responsibility.

So we might want to say that:

> *"If A is morally responsible for an event e, then A performed an action or omission that was part of the explanation of e".*

This is one destination of my argument in this essay. I want to argue that talk of explanation, and not causation, is the better way to understand moral responsibility for doing and allowing. harm. In general terms, I want to argue that an account which features *explanation* is more instructive than simply focussing on causation.

Secondly, I want to show that one instructive way of thinking about explanation is the account offered by Aristotle in his *Physics* and deploy that account in developing my view.

Here is how my argument will proceed. In Section 2, I will try and differentiate between causation and explanation. In Section 3, I show that causation is problematic as a basis for moral responsibility and that explanation is a more useful concept. Once I have established this, I look in Section 4 to the account of explanation offered by Aristotle in the *Physics*, stressing that what is often called the '*Four causes'* is best understood as the '*Four explanations'* or possibly the '*Four becauses'*. In Section 5, I argue that focussing on Aristotle's four-fold account of explanation produces a much richer understanding of moral responsibility than merely focussing on causation. In particular, it allows one to make a link between explanation and intention.

4.2 What is the difference between causation and explanation?

In this section, I hope to show that explanation is a broader concept that causation - and has much wider resources for attributing moral responsibility that causation and causal explanation.

Causation, plausibly, seems to be a matter of how things come to be, how certain things follow on from one another. As Peter Strawson argued (Strawson (1992, p115):

> *Causation is a natural relation which holds in the natural world between particular events or circumstances*[62]*.*

[62] In making this claim, Strawson is perhaps being Humean in his understanding of 'nature' and 'natural' than other philosophers, for example, Aristotle

The relata of causation are things (events or circumstances, on the view above). In that sense, it is similar to temporal succession (one thing following another thing) or spatial proximity (one thing close to another thing). One clue as to when we are talking about causation is the usage of the word '*caused*'. Examples might include:

- Suzie *caused* Tom to fall off the cliff
- Greenhouse gases *cause* global warming
- The lion *caused* the antelope to run away
- Lack of sleep *caused* John to be irritable

All these are things that happen in the natural world. The output of causation (Tom falling off the cliff, the earth's global warming, a running antelope, John's irritability) are all things in the world.

Carolina Sartorio puts this well: (Sartorio 2022, p 64)

> *First of all, I must note that, as a metaphysician, the concept of a cause which I am interested in is a natural concept. That is, it is a concept that picks out a certain relation between events or states of affairs, one that exists 'out there in the world'....it is not itself a normative concept. In particular, being causally responsible doesn't require being morally responsible for something. Natural phenomena such as tornadoes are causally responsible for outcomes, but they are not morally responsible for anything. The same goes for moral agents like us: we can be causally responsible for outcomes without being in anyway morally responsible for them as when we trip over someone accidentally and faultlessly cause them harm.....*[63]

Similarly, Beebee (2004, p296) says:

[63] Nothing new here – I've argued many times already in this dissertation that whilst causation might be necessary for moral responsibility, it is not sufficient for moral responsibility. Later, I show that causation is not even necessary for moral responsibility

> *But nobody within the tradition of metaphysics of causation that I'm concerned with here thinks that causal facts depend on human dependent norms.*

Beebee thinks this because if causal facts were dependent on human dependent norms, then this would make the truth of causal claims very relative.

By contrast, unlike causation, explanation is *not* a relationship that holds in the natural world. It is an intellectual or rational relationship. Explanation holds between facts or truths, but not between things in the world. An explanation is an account of the structures of patterns by which we render events intelligible.

When talking about explanation we do not use the word *cause* (as in causation), but words such as *because* or *due to* or *the reason why*.

Thus, another way of thinking about explanation is as a way of providing one possible answer to a '*Why*?' question[64].

By asking a 'Why?' question, the explainee suggests that there is a gap in their understanding – a gap that can be filled by grasping an explanation.

Explanations vary in quality - they can be wrong, or muddled, or long winded, or inappropriate or irrelevant.

Explanations are often contrastive – so an explanation of why you visited Oxford ('to see the cathedral') may include an explanation of why you did not go to Cambridge (because Cambridge has many fine buildings, but no cathedral).

The output of an explanation is not an event in the world (as in causation), but some form of understanding or perhaps some kind of knowledge.

[64] Famously introduced by Aristotle in *Physics* II. 3 – 'For the point of our investigation is to acquire knowledge, and a prerequisite for knowing anything is understanding why it is as it is'

Examples of explanations (these are actually all *causal* explanations, but explanations are not all of this sort) might include:

- *Why did Tom fall off the cliff? – he fell of the cliff because Suzie pushed him*
- *Why is the world warming? - because of the increase in greenhouse gases*
- *Why did the antelope run away? - because the lion chased it*
- *Why was John irritable - because he was short of sleep*

Other types of explanation than causal ones exist. For example, these are all explanations, but are most certainly not about causation

- Why can thirteen strawberries not be divided equally between three children? (without dividing individual strawberries) - because thirteen is not divisible by three (Lange 2013),
- Why are Socrates and this bust of Socrates similar? - because they both share the property of being snub-nosed (Brenner 2021)
- For example, why can we not make a circuit of the bridges of Konigsberg that involves crossing each of the bridges exactly once? – because of the topographical arrangement of the rivers and the bridges (Woodward and Ross 2021)
- Why do previously high blood pressures tend to settle towards the mean with repeated measurement? – because of a statistical phenomenon known as regression to the mean (Barnett, Van De Pols and Dobson, 2005).

A successful explanation has many purposes. It can help the explainee make future predictions and inferences. It can provide information and reasons for problems with the system. It can form the basis for praise or blame. It can justify future actions. It

can form the basis for strategies to alleviate problems. It can provide pleasure from the pure understanding[65].

I argue that causal explanation is a broader concept than causation. As we have seen, many explanations are causal in nature. But not all causal explanations involve causation; there are other ways to provide a causal explanation, such as by mentioning (for example) the role of omissions, or by ruling out alternative hypotheses.

There are also many types of explanation that do not offer a story of causation – we have seen earlier that there can be mathematical explanations, metaphysical explanations, topological explanations.

But it seems that to offer an account of causation is always to offer some kind of explanation – or at least, all accounts of causation can subsequently be used as some kind of an explanation.

In summary, then, it seems that causal explanation is a type of explanation (an explanation based on its role in causation) but it is not the only mode of explanation. I am going to show that in order to understand the difference between doing and allowing harm, we need to move away from relying on a sole means of explanation and widen this out to several means of explanation.

4.3 Causation v explanation in grounding moral responsibility.

As outlined in the previous chapter, the link between causation and moral responsibility is intuitive and plausible. The claim that causation grounds moral

[65] Rational humans have a desire to understand the world, and when we understand successful explanations, we fulfil that desire.

responsibility is widespread in the literature. Carolina Sartorio (2007) makes a typical responsibility claim:

> *Moreover, when an agent is morally responsible for an outcome, the existence of causal responsibility partly explains the existence of moral responsibility: the agent is morally responsible, in part, because he is causally responsible.*

The idea here is that the causation is necessary for the moral responsibility; without the causation there is no moral responsibility.

I want to challenge this claim. There seem to be at least three types of situations where there is clearly moral responsibility, but without any causal responsibility. However, in all three situations, there is an explanation available that could ground moral responsibility. So my suggestion is that these cases not only serve as counterexamples to Sartorio's view, but they also provide support for my suggestion that it is the existence of relevant explanatory links that grounds the attribution of moral responsibility.

The three situations are – omissions, over-determination, and causal selection.

a) Omissions

I have discussed this at length earlier, so will be brief. It seems incontrovertible that agents can be responsible for their omissions, yet if (as many philosophers claim) omissions do not (cannot) cause anything, then it is difficult to see how moral responsibility for omissions can be grounded in causation.

Example – the gardener who fails to water the plants is responsible for their death, even though (arguably) he has not caused their death. What we can say though is that the gardener's failure to water *explains* both the plant death and the gardener's moral responsibility for the plant death. We say that the plants died *because* of the

gardener's failure to water, and we say the gardener is responsible *because* he failed to do the watering. So here moral responsibility seems in fact to be grounded not in causal power of omissions (which may not actually exist), but in their *explanatory* power. The lack of watering explains the death of the plant, and answers the question – Why did the plants die?

b) Overdetermination

Imagine a firing squad of six members – each member fires a bullet all six bullets hit the victim simultaneously - and the victim dies. Each bullet would have been enough to kill the victim on its own. This produces puzzling intuitions about causation – is the action of each of the shooters an individual cause of the victim dying? Or do we say that *someone* caused him to die? Or is it some kind of collective causation event? (but how can there be causation without individual events causing the outcome?). In cases of overdetermination there does seem to be a disconnection between responsibility and causation – if the firer had not fired, then the victim would still have died.

However, despite uncertainty about whether there is causation or not, we would still want to say that someone is responsible for the death. And once again explanation seems to work better as a source of justification for responsibility. Although we cannot say that each shooter *caused* the death of the victim, we can say that each shooter's action *explains* the death.

And so once again the existence of a relevant explanatory link seems seems a better candidate to ground the moral responsibility of the shooters in a way that a mere appeal to causation cannot do.

c) Differentiating cause and background conditions

Now consider another case where it hard to ground moral responsibility in causation. Suppose a car swerves onto the pavement and injures a pedestrian. Perhaps the driver is late for a meeting, whereas the pedestrian is simply walking along. (Broadbent 2013). Intuitively, we would say that the driver *caused* the injury by swerving off the road and is therefore morally responsible for the injury. Indeed, the law courts employ a (counterfactual) test for causation, the *sine qua non* or *but for* test. *But for* the driver swerving off the road there and then, the injury would not have occurred. Based on this test, the driver might be found guilty and face a financial penalty or even a custodial sentence.

But this reliance on causation is problematic. The pedestrian also satisfies the *but for* test. *But for* the pedestrian being present at the moment of impact, the injuries would not have occurred. The pedestrian caused the accident just as much as the driver did. The driver might go on to argue that one of the causes of the crash is being prioritised over other causes. If causation is the criterion for moral responsibility, then both driver and pedestrian seem to share morally responsibility. This is clearly wrong – the driver but not the pedestrian is responsible, but this difference is not reflected in a causal story. One way of preserving this moral intuition about responsibility is to make use of the necessary/sufficient difference in ascribing moral responsibility – both the driver and the pedestrian were *necessary* for the crash, but only the driver's behaviour was *sufficient* to ground moral responsibility for the pedestrian's injuries. So one could tell a story about voluntary action and control and negligence and so on. But I think this describes the accident incorrectly. It seems that we do not want to say that the pedestrian caused the accident *at all* – but this is difficult to achieve within our understanding of the term

causation.[66] What seems better is that the behaviour (rushing, not concentrating etc) of the driver *explains* the accident and also explains the subsequent responsibility in a way that causation does not. The key to this explanation is no causal. It is that the driver had a blameworthy goal, whilst the pedestrian did not. In fact, importantly, *nothing* in the pedestrian's behaviour plays a relevantly similar explanatory role.

So in all three of these situations (omissions, overdetermination, causal selection), it seems that the link between causation and moral responsibility is not as robust as it seemed earlier. Can replacing 'causation' with 'explanation' do better?

An appeal to some relevant kind of explanatory link certainly seems to cope better with the above counterexamples than an appeal to causation alone. There are two problems with the appeal to explanation however – at present we have no real theory of explanation, and so our appeals to the significance of explanation are too vague (exactly what explanation are we offering to do the work of justifying responsibility claims?).

In addition, explanations just don't seem to be always the right sort of thing to ground moral responsibility – consider the examples given earlier involving mathematical explanation (dividing up the strawberries) and metaphysical explanation (Socrates and his snub nose). How can this be related to moral responsibility?

[66] It might be argued that the counterfactual ('but for') account of causation used by the courts is not the right account of causation for ascribing moral responsibility in cases like the one described in the text. But using (for example) a production account of causation seems to face the same difficulty – that both the driver and the pedestrian are causally responsible for the accident.

What we need is a theory of explanation that is closely linked to responsibility and also avoids the disadvantages of trying to ground responsibility in causation. Such a theory (I argue) is available – Aristotle's '*Four Causes'*.

4.4 Aristotle – explanation not causation – four causes are actually four 'becauses'

One established account of explanation comes from Aristotle in Physics II (3) – the so-called 'Four Causes' (Aristotle, trans Waterfield 2008). This account appears initially unhelpful – it seems to not be about explanation at all, as on first glance it appears to be all about causation (the section heading in the Oxford translation of the *Physics* is 'Four types of *Cause'*). However, I will claim (following a number of reputable interpreters[67]) that the 'Four Causes' is actually an account of explanation rather than causation.

To summarise, Aristotle claims that to give an account of change in a natural substance, it is usually[68] necessary to cite all four causes. The Four Causes are *material* causes, *formal* causes, *efficient* (agent-patient) causes, and *teleological* (final) causes. *Material* cause involves an account of a change which refers to that from which a thing is made – for example the bronze of a statue or the silver of a bowl. *Formal* cause involves an account of a change which refers to the form or pattern or essence of something, such as the shape of a bowl. This is quite a broad concept – Aristotle mentions, for example, the form of an octave as being a ratio of

[67] For example, Annas (1982), Bogen (1974), Coope (2007), Cooper (2013), Hocutt (1974), Vlastos (1969), Hennig (2009).
[68] Usually, but not always. For example, the explanation of an eclipse does not involve a teleological explanation

2:1 in frequency. The *efficient* cause involves an account of a change which appeals to the source or agent of the change or rest. Aristotle gives the example of the deviser of plan, or a father as a cause of a child, or a changer as a cause of a change. (This concept of cause is close to, but not identical to, the widespread Humean account of cause[69]). The *final* (or teleological) cause refers to what something is for. It involves an account of a change which appeals to the telos or goal – for example 'walking is for health'. Final causes can be characterised by the 'in order to' clause – one walks (for example) *in order to* be healthy.

Aristotle suggested that this was an exclusive list of causation – '*these are more or less all the ways that we use the word cause.....all so called causes fall under these headings'*

Aristotle uses the word 'cause' many times in this section – so why might it be thought that he is talking not about causation but about explanation? I think there are four reasons why this is a plausible interpretation of his view.

Firstly Aristotle begins the section by saying - "*For the point of our investigation is to acquire knowledge, and a prerequisite for knowing anything is understanding why it is as it is – in other words grasping its primary cause*".

So the 'Why' question seems fundamental to Aristotle's argument. So it seems natural to see explanation rather than causation as a response to a 'Why' question – causation can sometimes provide some kind of an answer to a 'Why' question, but not always.

Secondly, and linked to the above, the Greek word *aitiai* used by Aristotle in the *Physics* can be translated as both 'cause' or 'explanation'. So there does not seem

[69] I'm not assuming here that the Humean account of causation is necessarily correct – for example, Hume's claim that causation must precede effect can be challenged – as can Hume's claim that causes must be spatially adjacent to effects.

to be any textual reason for preferring cause over explanation. If we think of *aitiai* in terms of answering a 'Why' question, then it seems plausible to think of the Four Explanations rather than Four Causes.

Thirdly, the 'causes' that Aristotle discusses are fundamentally different from the default (broadly Humean) concept of cause that I have been discussing in earlier chapters. To speak of Aristotelian *causes,* risks misleading from the beginning. Humean causes capture only one sort of Aristotle's Four Causes (the efficient cause) and even then, it is not clear that Aristotelian 'agent-patient' causation is the same as our modern understanding of (so called) efficient causation,

Finally, many of the uses to which Aristotle puts the word *cause* seems counter-intuitive unless the word is interpreted as an explanation. The best example of this is perhaps teleological 'causation'. Even to think about teleological causation is difficult if one thinks that the telos is literally supposed to be 'causing' the explanandum to occur - as it suggests that the end to which one aims towards somehow brings about a change that precedes the aim: one runs the risk of visualising teleological explanation (to quote Julia Annas's memorable phrase) as '*ghostly tugs from the future'*. (Annas, 1982). It makes much more sense to think of teleological *explanation*, where talks of ends or purpose answers a 'Why?' question about a particular activity.

Vlastos (1969) and Hennig (2009) make similar points - they argue that Aristotle's talk of four causes is incomprehensible if we understand cause in terms of production (that is, efficient causation). So it makes no sense to say that the bronze *produces* the statute's width, that the abstract entity equal to two right angles *produces* the rectangularity of the semicircle, and the health which does not yet exist *produces* the walking which does exist.

They both argue that we understand Aristotle's Four Causes better when we translate 'X is the *aitia* of Y' as – '*Y happens or is the case because of X*". (see for example, Hennig (2009) p293). For example – why is the statue heavy? – because it is made of bronze – why is the man walking? – because he wants to be healthy. William Charlton (1992) is unequivocal when he says– *'the discussion of aitia is a discussion of explanation, and the doctrine of four causes is an attempt to distinguish and classify different kinds of explanation".*

Moravcsik is equally committed to the idea that an aitiai '*is simply an explanatory factor, whatever this may be....it is quite misleading to refer to the theory of aitiai as the doctrine of causes. It is rather a theory about the structure of explanation*". Moravcsik (1974, p3)

To conclude this section therefore, we see that we have established what Aristotle is referring to when he uses the term aitia or cause. He is talking about explanation, and not what we call efficient causation. Furthermore, he suggests that explanation comes in four types, not just what we tend to understand today as (efficient) causal explanations but also as explanations related to form, to material and to telos.

4.5 Making use of Aristotle's account of explanation

If we accept the above argument, that Aristotle's Four 'Causes' are really 'Four modes of explanations" (or perhaps 'Four Becauses' (Hocutt 1974) - does this help with providing (firstly) an overall explanation for events or processes, and secondly an attribution of moral responsibility for events and processes? I think it does, in the following way.

Consider the speeding driver, driving too fast in order to get to work on time. The car swerves and hits a pedestrian who just happens to be walking along the pavement in the vicinity of the speeding car. The car mounts the pavement, collides with the pedestrian, and the pedestrian suffers several broken bones and bruises.

Firstly, what is the explanation of the pedestrian's injuries? This can be explained in the following way.

Material explanation– the car is made of hard, non-flexible material, whereas the pedestrian is composed of soft tissue, and harder, but brittle bone.

Formal explanation – the car is travelling at speed, going too fast for the conditions, and the interface between driver and car is insufficiently adaptable to allow the driver to steer away from the pedestrian.

Efficient explanation – there is a process of exchange of momentum, such that the pedestrian is propelled into the air, and lands painfully on the ground. In more detail, the agent (the car being driven as it is) acts on the pedestrian (the patient) so as to knock him over. In Aristotle's terms, to actualise his potential to be knocked over.

Final explanation – the driver's aim was to get to work on time, despite leaving insufficient time for the journey. He did not have an aim (at that moment) of driving safely, his only aim was to avoid being late for work. The way that the driver is driving is very relevant to his moral responsibility for the pedestrian's injuries – his aim is to get to work on time, irrespective of the risk to others (a risk that was actualised in this accident).

By contrast, the pedestrian's aim was to enjoy a leisurely stroll along a safe footpath. The pedestrian did not have an aim that was relevant to moral responsibility- his telos is morally neutral, and hence he is not morally responsible for the accident – even though in purely causal terms he contributed to the accident.

All four modes of explanation play some role in their being an issue of moral responsibility, and in determining where the moral responsibility lies. Importantly, had any of the four explanations been different, the pedestrian would not have been injured – or perhaps not injured to the same degree. Had the car been made of soft material, or been travelling slowly, or had missed the pedestrian, then no harm would have been done. Most importantly, had the driver's aim not been to get to work on time (which involved driving too fast) then the accident would not have occurred. What about moral responsibility for the accident? Does Aristotle's four-fold system of explanation provide any insights into the correct attribution of moral responsibility for the accident? If we accept that the behaviour of the driver *explains* the accident (In Aristotle's four-fold schema) then it seems that, in the absence of exculpating factors (see below) the driver is morally responsible for the accident. The driver could have chosen a safer car with different material (softer outer shell) or different form (less angular). The driver could have driven slower and ameliorated the efficient explanation of both the accident and the pedestrian's injuries. Most importantly, the driver could have had a different telos – his aim could have been to drive safely rather than to get to work on time. It is this mode of explanation that takes us to the attribution of moral responsibility.

Aristotle seems to prioritise the *telos* of the driver above the other three explanations for an outcome. In this story the telos of the driver was inappropriate for someone driving a potentially dangerous vehicle at speed in an area where pedestrians co-exist. Consequently the driver is morally responsible for the accident. He would have also been morally responsible if his telos has been to cause harm to that particular pedestrian.

To summarise Aristotle's account of the attribution of moral responsibility for harm, this seems to rest on whether an action or behaviour of the agent explains the harm.

4.6 Conclusion

In this chapter, I have tried to argue that the causation entailment claim suffers from counter-examples – situations where an agent is morally responsible for an outcome even when it is uncertain as to whether he has caused that outcome. The explanation entailment claim suggests that an agent will be morally responsible for an outcome if his action or behaviour is part of the explanation of that outcome – and that the relevant mode of explanation does not have to rely on an account of causation. Furthermore, in Aristotle's account of the 'Four causes' (actually 'Four Explanations') there is a plausible account of how explanation works, and how this can link to moral responsibility.

In the next chapter, I will say more about one particular type of explanation, the teleological explanation, which has particular relevance to the attribution of moral responsibility for both doings and allowings.

CHAPTER FIVE – INTENTION AND TELEOLOGICAL EXPLANATION

5.1. Introduction

Where have we got to so far? I have shown (Chapter 3) that there is some difference in kind between doing harm and allowing harm, and this difference is related to the difference between genuine causation and causal explanation. Doing harm genuinely *causes* harm (and at the same time provides a causal explanation of the harm) whereas allowing harm does not cause harm, but does provide a causal explanation of the harm.

To summarise:

Doing harm – genuine causation of harm (which is one way of providing a causal explanation)

Allowing harm – a causal explanation only (no genuine causation)

And, I suggest, this can provide an explanation of the moral difference between doing harm and allowing harm, based on the moral implications of the genuine causation/causal explanation difference. Doings genuinely cause; allowings only causally explain – and plausibly it is morally worse to genuinely cause a harm, rather than just being part of a causal explanation of that harm. So in general, and as a first approximation, doing harm is morally worse than allowing harm.

However, this can only be part of the story. The genuine causation v causal explanation story can explain the moral difference between doing and allowing harm, in the way suggested in Chapter 3 of this thesis. But it cannot explain the moral difference between different sorts of allowing – for example, the difference in moral responsibility between the gardener (who allows the plants in his care to die), and (for example) the Pope – whose non-watering can also provide a causal explanation of the death of the plants - but who is in no way responsible for their death.

Nor can it explain the fact that there is sometimes a *lack of difference* between the wrongness of doing harm and allowing harm – for example in Rachels' (1975) Bathtub case. In this case, it seems that the action of Jones (who allows the child to drown) is *every bit* as egregious as the action of Smith (who pushes the child under the water). However, the genuine causation v causal explanation account would have it that what Smith does is morally worse than what Jones allows – and this does not seem to be the case.

It therefore seems that the genuine causation v causal explanation explains some of the differences and similarities in the wrongness of doing and allowing harm – but does not offer us a complete moral picture. In particular it does not provide a good explanation of the varying wrongness or blameworthiness of different types of allowing of harm.

In this chapter, I claim that the *intention* of the agent is important for the rightness and wrongness of both doing and allowing harm, and this affects the attribution of moral responsibility for these two behaviours. I claim that it is only by providing an intention-based or teleological explanation can we fully understand moral responsibility for doing and allowing harm. In particular, an intention-based

explanation allows us to understand why *some* allowings of harm are wrong, whereas *other* allowings of harm do not carry that moral burden. This leads me to develop an account of doing and allowing harm that might be called the *coarse-distinction/fine-distinction* account. I argue that whether a behaviour genuinely causes or merely provides a causal explanation – gives us only a *partial* understanding of the morality of the particular behaviour. Such an evaluation is incomplete – a teleological explanation adds fine detail to the account, and allows fine distinctions to be made. In particular, the fine distinctions that I am interested in relate to the variation in the blameworthiness of different doings and allowings.

In the rest of this chapter I will explain what I mean by teleological explanation, and what I mean by fine and coarse distinctions.

This chapter is in several sections.

Firstly (section 2) I introduce the idea of intention, relying mainly on the account of intention offered by Elizabeth Anscombe in her monograph, *Intention* (1963).

Next (section 3), I make a claim that intention matters in the attribution of moral responsibility and rightness and wrongness for both doing and allowing harm. This claim starts with a commonsense argument, based on situations where intention is or is not present. However, this intuition-based claim is rather weak, and I need to provide further evidence for my claim that intention matters for the rightness or wrongness of doing and allowing harm. In particular, I need to defend this claim from several opposing philosophical assertions which claim that intention is *irrelevant* to the attribution of moral responsibility and to the rightness and wrongness of actions.

I then go on (section 4) to show *why* intention matters for the attribution of moral responsibility. I argue that intention is important for two reasons.

Firstly, because an intention can provide explanatory information about a particular action and hence allow one to understand the rightness or wrongness of that action. I argue that one cannot fully understand what doings and allowings are, unless one understands the intention and aims which motivate these behaviours.

Secondly, by understanding intention and the related concept of a teleological sequence, it allows one to link the intentional doing or allowing to moral considerations.

In particular, (section 5) I argue that moral responsibility for both doings and allowings rely in some way on intention; in other words I claim moral responsibility for doing and allowing harm involve a teleological explanation.

In the final section of this chapter (section 6) I introduce the concept of fine distinctions and coarse distinctions - and show how causal factors and intention-related factors interact.

This will allow me to plug some explanatory gaps that I highlighted at the beginning of this chapter.

What are these explanatory gaps? In summary, these are (firstly) to be able to provide a comprehensive explanation of why agents might be morally responsible for both the doing and allowing of harm. I argue that a combined causal and teleological account can do this. In particular, a combined causal/teleological story can go some way towards explaining the variation in the moral difference between doing and allowing harm alluded to earlier in this thesis. It can explain why in some situations there is a moral difference between doing and allowing harm, whereas in other situations there is no difference.

In addition, I show that a teleological explanation is able to explain why two agents who both fail to do the very same thing (water the plants for example) can have widely differing blameworthiness for a bad outcome (the plants dying, say). Such an account can show why the gardener's failure to water *explains* the death of the plants, whereas the Pope's (and many other individuals') failure to water *cannot* explain this.

My claim is this; that by adding an account of intention-related (teleological) explanation to our existing account of causal explanation, we end up with a more comprehensive account of the differing degrees of moral responsibility for doing and allowing harm.

5.2 What is intention?

I have no wish to develop a detailed theory of intention (in itself enough for several PhD theses) but further discussion will be assisted by having some rough and ready notion of what intention is.

Intention is a complex notion, and Elizabeth Anscombe (1963) in her monograph *Intention*, famously thought about it in under 'three heads' (§1), which she tried to unify. I illustrate Anscombe's subdivision by reference to the act of intentionally prescribing an antibiotic to a patient, in order to cure their chest infection.

Firstly, there is intention when considered as an intentional action. This is intention as an adverb - I prescribed the treatment to the patient *intentionally*. In other words, I prescribed the antibiotics as a result of deliberation. My prescribing was for a reason, and not the outcome of some random process (I didn't for example, just toss a coin to discover which drug to prescribe).

Secondly, there is the forward-looking idea of intention or *intention for the future* - this is the idea of intention as expressed in a verb – 'to intend'. This seems to map onto a mental state. Example: I *intend* to prescribe the treatment in my afternoon clinic. I may not have started the prescribing process yet, but I intend to do it later today. This is defeasible: it may be the case that there is an emergency, and the treatment never gets prescribed.

Finally there is the idea of intention '*with which'* something is done. This is intention as a noun - I prescribed the treatment *with the intention of* making the patient get better. This can be equally well expressed in terms of the conjunction '*in order to'* – I prescribed the antibiotic '*in order to'* make the patient feel better.

To summarise the three 'heads' of intention:

- A is x-ing intentionally (Adverb) – 'I intentionally prescribed the medication'
- A intends to do y (Verb) – 'I intend to prescribe the medication'.
- A is x-ing with the intention of doing y (Noun) – *'My intention in prescribing the drug was to cure the patient'*

Anscombe argued that a satisfactory account of intention should be able to unify these three concepts of intention. She further argued that the way to unify these three concepts was by an understanding of the meaning of intention as '*in order to'*.

Fairly uncontroversially, there seems to be a link between intention and the agent's aims. We intend what we aim at. So, in differentiating intentional actions from non-intentional actions, we rely on the notion of aims, purposes, or goals. This is consistent with Anscombe's emphasis on the '*in order to'* concept.

For example, a recent monograph on intentional action offers this definition:

> *An agent intends to X when it is her aim, goal or purpose to X, she X-s intentionally when X-ing is among her aims, goals or purposes in acting, and*

> *she brings about an effect intentionally when it is her aim, goal or purpose to bring it about.* (Stuchlik, 2022, p82)

By contrast, and for completeness, here is a definition of what it means to act non-intentionally, or incidentally.

> *An agent X-s incidentally when X-ing is not amongst his aims, goals or purposes in acting, but she knows or is aware that she is X-ing; she brings about an effect incidentally when it is not among her purposes, but she is aware (knows, foresees, expects) that it will result, or likely result from her conduct*. (Stuchlik, 2022, p82)

Here is an example that clarifies this difference.

Consider again the patient with a chest infection, who is prescribed antibiotics. A non-intentional, foreseen consequence of this action is that the patient develops a rash. I don't intend the patient to develop a rash (giving the patient a rash is not my aim), but it is a recognised side-effect, and I can foresee that it might occur. (There might also exist non-intentional, completely unforeseen consequences of intentionally prescribing the antibiotic, but I will not discuss these here).

Understanding the intention of an agent helps us to successfully answer a 'Why?' question about a particular action. '*Why did you prescribe the antibiotic*?' – produces the answer '*In order to cure the patient of the infection'*. The intention of the agent reveals the reason why the agent acted in that way. The answer to the 'Why?' question is not – '*To give the patient a rash'*.

5.3. Intention matters for the assessment of rightness and wrongness

After clarifying what intention is, I now want to suggest that intention is important in the attribution of blame for actions. This is in contra-distinction to one popular philosophical view that claims intention is *not* relevant to moral responsibility.

First – a preliminary intuitive claim. That intention has *something* to do with moral evaluation does seem very plausible. Here is an example (originally from Phillipa Foot (1967, p2) and quoted in Stuchlik (2022, p24).

> **Scapegoat**. *A person could kill an innocent man and pass him off as a criminal in order to prevent a mob of rioters from killing five hostages.*

> **Trolley driver**. *The driver of an out-of-control trolley will run down and kill five workmen unless he steers the trolley onto a sidetrack. There is one (innocent) work man on the sidetrack whom the driver will kill if he turns onto it.*

It seems uncontroversial that it is permissible to steer on to the side track in Trolley Driver, but impermissible (in fact morally outrageous) to kill the innocent man in order to satisfy the mob.[70] Yet the outcome is the same (one innocent man killed, five saved) in both scenarios. Why the difference? One plausible explanation for the difference is that in *Scapegoat* the innocent man is killed intentionally, whereas in

[70] Many other cases are available that make the same point. The most famous of these (for example, Stuchlik (2022) p24) is *Transplant*, where an innocent man who is attending the hospital for a check-up can be killed in order to harvest his organs that would go on to save five others with organ failure. As in *Scapegoat*, but in contrast to the situation in *Trolley Driver*, it seems impermissible to kill the hospital clinic attendee – because the killing would be an *intentional* killing.

Trolley Driver, the innocent man is killed incidentally – the aim (=intention) was to save the five work men, and the killing of the one workman is a foreseen, but completely unintended consequence of this aim (had the trolley missed the one work man, the driver would not feel he had failed to achieve his aim – in fact he would be delighted).

However, this claim about the moral implications of intention is not universally accepted. There is an opposing view that suggests that it is inappropriate to attribute blame to agents who have the intention to do harm, but yet do not actually translate the intention into harm (what might be called 'pure' intending).

Examples of this opposing philosophical opinion include James Rachels (1986), Gerhard Overland (2014), and Judith Jarvis Thompson (1991). This view regards intentions to be *irrelevant* to the question of moral permissibility. So my first task in this chapter is to argue against this collections of arguments – which might be collectively called '*the irrelevance theory'* (Stuchlik (2022), p134). There are several versions of this argument, but they all rely on an underlying mistake about the locus of intention. Exploring the underlying mistake will provide valuable insight into the role of intention.

5.4 Intention does not matter for the assessment of rightness and wrongness of actions – Overland's 'all in the mind' argument – and an Anscombian response.

Overland (2014, p482) and others, wonder why what goes on in the mind of an agent should be relevant in determining whether killing a person is permissible. Overland argues that whether the agent does something incidentally or intentionally is a fact about the agent's state of mind rather than a fact about the world. As he puts it (p

482) – '*it is not clear why the constraint against killing innocent people should be affected by one's not intending their death'.*

The idea here is that we cannot be held morally accountable for activity that goes on *just* in our mind. After all, we often have brief mental thoughts (which could be easily considered to be intentions) that just appear from nowhere and it seems wrong to hold us responsible for these sorts of thoughts. By contrast we can certainly be held blameworthy for things that we bring about in the world. But whether we bring about things intentionally (or not) isn't a claim about the state of the world; it is (so the argument goes) a claim about our state of our mind and this is not a proper subject for blameworthiness. To quote Stuchlik (p137) – *'aren't thoughts one thing and actions another'?'*

Response to Overland's argument

I argue that Overland's '*all in the mind'* objection does not hit home. The fact that an agent acts intentionally is not *just* a fact about his or her state of mind. It is also a fact about what is going on in the world. This mistake was identified by Elizabeth Anscombe in her *Intention* (1963).

Anscombe wants to deny (as do I) an understanding of intention whereby intention is *only* a mental state. She thinks that this particular usage of the word leads to moral confusion. This usage is as follows (Anscombe (1961, §27)

> 'According to this [Cartesian] psychology, an intention was an interior act of the mind which could be produced at will. Now if intention is all important – as it is – in determining the goodness or badness of an action, then, on this theory of what intention is, a marvellous way offered itself of making any action lawful. You only had to 'direct your intention' in a suitable way. In

> practice, this means making a little speech to oneself: “What I mean to be doing is....”

Anscombe feels that this kind of usage of the word ‘intention’ is illegitimate.

She makes a similar argument at two more points in *Intention*

> ‘A man’s intention is not so private and interior a thing that he has absolute authority in saying what it is – as he has absolute authority in saying what he dreamt’ (§22):

and

> ‘If an intention is an interior movement it would appear that we can choose to have a certain intention and not another by just, for example, saying within ourselves: ‘What I mean to be doing is [X] and not [Y]. The idea that one can determine one’s intention by making such a little speech to oneself is obvious bosh”. (§25)

Anscombe called this approach a ‘perverse doctrine’. She did not see intention as ‘just’ a mental activity – rather, intentions are part of what she called a ‘teleological sequence’ which was determined by the aim of the doing.

She saw intentional action in terms of behaviour for which it is meaningful to raise the question ‘*Why’* in the sense of enquiring about reasons for the action. To use Anscombe’s words (§5):

> *‘what distinguishes actions which are intentional from those which are not?.... they are the actions to which a certain sense of the question ‘Why?” are given application; the sense is of course, that in which the answer, if positive, gives a reason for acting’.*

Furthermore, Anscombe thought that the same action could be both intentional and non-intentional – intentional under one description and non-intentional under another description.

Anscombe's example is *sawing* - this action is intentional under the description '*sawing a plank'* (so it is possible to ask the question – '*Why are you sawing a plan*k?' under that description). Positive answers that would give a reason for the action (and therefore would reveal the agent's intention) might be '*To build a fence'* or '*To make a table*'. It would not be intentional under the description - '*making a squeaky noise'*. Why? - because the agent's reasons for performing this action have nothing to do with noisemaking and everything to do with (say) building or making.

To go back to my example (in the case quoted above), if I was asked '*Why did you prescribe the antibiotics?'* I can reply, '*In order to cure the chest infection'*. It doesn't make sense to answer the 'Why?' question with the statement: '*In order to produce a rash'* as this was never my intention. Anscombe pointed out that intentional action can only be understood '*under a description'* - my intention was to give the antibiotics under the description of '*curing the infection*' but it was not my intention to give antibiotics under the description of '*producing a rash'* (because producing a rash was not the reason I prescribed the antibiotics – and was not why I acted in this way).

This idea is linked to the idea of planning – when an action is planned the agent is able to spell out the reasons why an action is done – in this case to cure the infection (and not to produce a rash) – or (in the case of the sawer of wood) 'to build a table'.

Anscombe saw intentional action as action which has (what she called) a *teleological order.*

What makes intentional action intentional? As discussed earlier, intentional action (thought Anscombe) is action where a certain sense of the question 'Why?' has application[71]. (§3). The relevant sense of the question is that the answer (if positive), gives a reason for acting. Consider the question – '*Why are you climbing the stairs*?' This sort of question certainly has application. Anscombe thought that the best way to answer this sort of 'Why' question is to mention the telos or good which it is aimed at (§17). This may be an action in the future ('*I'm going to find my textbook'*) or an action that the agent is currently doing ('*I'm exercising my core muscles'*). Whether this answer cites the future or a current situation, the key point is that the question why can be answered by the phrase – 'in order to'. (*I'm going up the stairs in order to find my textbook'* or *I'm going up the stairs in order to exercise my core muscles'*).

It can now be seen that this understanding of intention is wedded firmly to the idea of teleology. There is an explanatory connection between the question '*Why'* and the answer '*in order to'*. The connection may explain why one action would lead to the other, or it may be more basic than this – the question may actually be constitutive of the answer. But the key commonality is that the answer given is always in terms of something desirable for the agent – the good. The good may be a derivative good (good for the sake of some further good) or a final good.

Furthermore, the '*Why'* question can be reiterated to form a teleological sequence. This can be illustrated by Anscombe's famous example of the man at the pump (§23). The man at the pump is moving his arms up and down, and in order to find out the man's intention, one asks a series of 'Why?' questions.

[71] Also intentional actions are actions that can be known without observation – but I will not consider this approach again

For example:

Q: Why are you moving your arm up and down?

- I'm operating the pump

Q: Why are you operating the pump?

- I'm replenishing the water supply to the house

Q: Why are you replenishing the water supply to the house?

- To poison the inhabitants

Anscombe says this teleological order is present whenever actions are done with intention (§23) and can be used to distinguish intentional from non-intentional (incidental) action. Actions that are not intentional are outside this teleological order. To take the example of the pumping man, the teleological order runs from arm movement to poisoning the inhabitants. When the agent is pumping, he might be making a clicking noise, or making himself sweat - but none of these actions are intentional. They are not intentional because making a clicking noise is not part of the teleological process that leads to poisoning the inhabitants - which is the agent's aim (and which the agent sees as good in some respect). It is merely a side effect of the pumping. In Anscombe's terms, the agent, when asked '*Why are you making a noise?'* - would reject application of the 'Why?' question with regard to the pumping action – so the noisemaking is not an intentional activity, but a merely incidental by-product of the pumping (though perhaps an incidental by-product that the pumper is prepared to tolerate in order to achieve the intentional aim of pumping).

It is also worth noting that there may be another teleological sequence for which the why question relating to the noise *does* accept application – for example the agent may intend to make a noise to amuse him which in turn may alleviate the boredom of all the pumping. This would then be an intentional action (under that description),

but as part of a completely different teleological sequence (and with a completely different aim).

It can thus be seen that teleological explanation and intentional explanation are the same kind of explanation. To explain by reference to intention is to explain teleologically – with respect to the agent's *telos*. And, it should be said, the *telos* for an agent is always some good or apparent good *to the agent*. Note that is not necessarily a moral good – it just needs to be a good *for that agent*. Some 'goods to the agent' can actually be morally bad – as in the cases of the evil cousins in Rachels' examples – who were prepared to kill in order to produce the good (to them) of financial reward.

Anscombe hence believed that intention was not just in the mind – the intention with which an agent acts is a calculation about what to do (§33) – such a calculation is about working out how to get to one's end. Such intention-based calculation starts with the thing aimed at (§34) which (according to Anscombe) is something that is desirable in some way. The agent also desires (instrumentally) what is a means to this (desirable) end.

So Overland's '*all in the mind'* argument does not work – intention takes place both in the mind but also in the world. In contrast to Overland's claim, it does not seem unreasonable to take intention into account when assessing the rightness or wrongness of an action. Indeed, we can only fully understand the rightness or wrongness of an action when we also understand the intention with which that action is done.

5.5. Intention does not matter for the assessment of rightness or wrongness of actions- Rachels' inconsistency argument (and a response from Audi).

James Rachels (1986, p93) in *The End of Life, Euthanasia and Morality* is also critical of the role of intention in moral assessability – he however presents his criticism in a slightly different way.

Consider the following examples, about two grandchildren visiting their grandmother.

> ***Jack.*** *Jack visits his sick and lonely grandmother and entertains her for the afternoon. He loves her and his only intention is to cheer her up. Jack does know that his visit might influence the making of the will in his favour, but that is not part of his plan.*
>
> ***Jill****. Jill also visits her grandmother and provides an afternoon's cheer. But Jill's only concern is that the old lady will soon be making her will: Jill wants to be included among the heirs.*

Rachels claims that (p93) '*Jack and Jill do the very same thing'* (they, according to Rachels, spent time with their grandma and cheered her up). Because '*they do the very same thing*', Rachels argues that there is no (and indeed, cannot be any) moral difference between their actions[72] – it is logically nonsensical to claim that what Jack did was right and what Jill did was wrong – because they do the very same thing. Jack and Jill both visit their grandmother for the afternoon. There is no difference in

[72] Though he does accept that Jack and Jill's character is different even though their actions are not – 'Jack seems admirable for what he did, while Jill does not'.

what they do. So how could we possibly justify assessing the morality of their actions differently? Hence (says Rachels) intention does not matter morally.[73]

Response to Rachels' argument

I want to argue that Rachels is mistaken – the action done by Jill and by Jack are not identical – they are not doing '*the very same thing'* as Rachels claims. I shall argue that when agents act (or, indeed, fail to act) with different intentions, they are actually doing different actions, with different moral imports. Furthermore, the difference the moral difference is not just about the agent's character (as Rachels avers) but about the rightness and wrongness of the action – and hence the blameworthiness of the individual.

I want to make this claim in a number of ways.

Firstly, in terms of specificity. Some act-types are more *specific* than others. Ralph Wedgwood (2011) makes a distinction within what might be called act descriptions. Wedgwood argued that one way to make act-types more specific is to include in the description of the act type an intention. Act-types without an intention are called by Wedgwood *thin act types*, and when the intention is added they are called by Wedgwood *thick act types.* To use a medical example:

> *Switching off a ventilator* is a thin act type, whereas
>
> *Switching off a ventilator with the intention of ending the life of a patient is a thick act type*

A thin act-type can only answer the question – '*What is the agent doing*?' – whereas a thick act type answers this question and answers the question – '*Why is he doing*

[73] Judith Jarvis Thompson (1991, p293) presents a similar argument to Rachels, but presents the inconsistency objection from the point of view of an impartial deliberating agent.

it?' The intention with which the agent acts (or fails to act) ''converts' a thin act type to a thick act type. Robert Audi (2016) makes a similar point. When we act (says Audi 2026, p51) we typically do things in a certain manner and with certain intentions. Audi uses the word '*conduct-type'* to describe doing something (or allowing something) in a certain manner and with a certain intention. (so Wedgwood's thick act-types correspond to Audi's conduct-types). Audi thinks that actions typically have several dimensions. Many doing words give information about several dimensions at the same time – the example Audi gives (p54) is 'yelling' – which tells the readers that it is speaking in a particularly loud or shrill way. Similarly (my example) murder is killing someone intentionally. St. Thomas Aquinas (1947) offers another example where the moral valence of the action is completely altered by a full appreciation of intention. Aquinas's example is this - giving alms is morally good - but giving alms in order to get praise is morally bad. The conduct type – giving alms to gain praise is the act that is wrong. The wrongness lies not in the giving, but in the vainglorious giving. By including the intention ('*in order to receive praise'*) it is possible to fully understand the action of giving alms. Without intention, we cannot understand what is going on. Note that the intention completely changes the morality of the action.

How do these considerations apply to the previously expressed objections?
If we think about the Rachels' objection, it can be seen that what Jack and Jill does are only the same if we consider the thin act type, not the thick act type. Or to use Audi's terminology, we can say that what Jack and Jill do are the same act type, but are different conduct types. Jack's act type and Jill's act type are the same – *'to spend a day with my grandmothe*r'. But Jack's conduct type is to *'spend a day with*

my grandmother in order to cheer her up' and Jill's conduct type is '*to spend a day with my grandmother in order to inherit some money'.*

Rachel's argument thus fails – the two grandchildren are *not* doing the same thing. It does not seem inconsistent to judge the two conducts differently. What Jack does and what Jill does are different actions, with different moral implications.

This argument actually links with Anscombe's ideas that were discussed in the previous section. Consider again the activity of the pumping man (Anscombe 1963 §23). Recall that the intention of the pumping man is presented by Anscombe as a sequence of questions and answers, as laid out below:

Q: Why are you moving your arm up and down?

- I'm operating the pump

Q: Why are you operating the pump

- I'm replenishing the water supply to the house

Q: Why are you replenishing the water supply to the house

- To poison the inhabitants

The thin act-type would be the initial '*moving your arm up and down'*. But this does not really tell us what is going on – this is only revealed by the thick act type (or, to use Audi's terminology) the *conduct-type* presented later in the sequence – moving one's arm up and down, in order to operate the pump, in order to replenish the water supply, *in order to poison the inhabitants*. Note alterations in the conduct type, whilst keeping the thin act type the same, alters the moral valence of the action completely. Consider this:

Q: Why are you moving your arm up and down?

- I'm operating the pump

Q: Why are you operating the pump?

- I'm replenishing the water supply to the house

Q: Why are you replenishing the water supply to the house?

- To refresh the inhabitants who are unwell with dehydration.

The thin act type in the first story is identical to the thin act type in the second story – but it is only when we understand the thick act type or the conduct type do we see that the pumping is a different action in the first sequence than in the second sequence. It is only when we that we are able to see the malevolence in the first story and the kindness in the second story that we are able to fully understand what the two actions actually *are*.

5.6 Preliminary conclusion

I have developed an argument in support of the importance of intention in the moral assessment of action. In summary, an understanding of the intention of the agent allows us to provide a full specification of the action – in other words, to fully understand what the agent is doing (or allowing). It also shows that intention is not just in the mind – it involves the world as well, and cannot be understood without understanding the teleological sequence within the world that explains why an agent acts. Anscombe had it right when she said (1963, §45) – '*Roughly speaking, a man intends to do what he does'.*

In developing this argument, I have slightly widened the concept of an action from the rather narrow notion of an instantiation of an act-type to the richer notion of act-conduct. When human beings act, they perform acts of certain types with certain intentions. The discussion so far in this chapter has concerned itself with these two

dimensions of human agency and I have argued (I hope successfully) that both are important[74].

Most of the examples given so far refer to doings, but it also seems true that intentions affect the moral valence of allowings as well as doing; if doing can be performed with bad intentions (as in the case of Jill, quoted above) so can allowings. Consider this case. Imagine a doctor allowing a patient to die by (say) not offering a life-saving treatment.

In version (1) of the story (*Allowing (1*), the doctor's intention is to spare the patient the suffering that would surely follow.

In version (2) of the story (*Allowing 2*) the doctor's intention was to inherit a sum of money, perhaps bequeathed to the doctor in recognition of previous excellent care.

The intention is not the same in Allowing (1) and Allowing (2). Just as with doings, the moral valence of the allowing is altered by altering the intention.

This completes my argument for the importance of intention in developing a complete understanding of moral responsibility for both doing harm and allowing harm.

It does seem beyond question that agents can be morally responsible for the harm they intentionally[75] allow (think of a gardener who intentionally omits to water the plants under his care, or a doctor who fails to treat a patient with a serious but treatable condition).

---

[74] Act-conduct as well as including actions and intentions also include the manner in which the action is done. I don't propose to discuss this feature of actions in this thesis, but might include (Stuchlik 2022, p61) – "*Suppose Charles disciplines his misbehaving child in a way that is excessively harsh. If discipline is merited by the child's behaviour, then disciplining her is the right thing for Charles to do. Still the fact that Charles disciplines the child in such a harsh manner surely means that his conduct is morally defective'*

[75] Of course, agents are also responsible for harm *incidentally* allowed, but the mechanism of the blameworthiness is different as discussed later in this chapter

Up till now, I have been claiming that the moral responsibility for allowing harm is based on its role in causal explanation. This view is plausible, and I think partially correct. I do not, however, think it is the complete answer.

Here is the problem. Let's go back to the gardener and the Pope again. I want to claim that the gardener causally explains (as suggested by Helen Beebee (2004, p307) the moral responsibility for the death of the plants in the counterfactual way discussed earlier. But then, worryingly, this includes too much – if the gardener is morally responsible for the death of the plants, then so is the Pope, whose lack of intervention also causally explains the plants' demise. And we clearly don't want to blame the Pope for the death of the plants – that would be absurd. But, reversing the argument, if we don't want to blame the Pope then we cannot blame the gardener either - which is also absurd (of course he is responsible).

So we seem to have reached an impasse. It seems that the causal explanation story for the moral responsibility for the death of the plants is plausible but cannot be completely correct - something is missing.

Sarah McGrath realised this was a problem. She spelt out the problem as below: (McGrath (2005, p131) whilst commenting on Beebee's notion of causal explanation.

> *Another difficulty for the view is that Barry [who had promised to water the plants], and not Carlos [who had not promised to water the plants, and in fact lives many miles away], is responsible for the death. Causation and responsibility are linked, albeit in some hard to spell out way; intuitively the fact that Barry, and not Carlos, is responsible is linked to the fact that Barry's not watering, and not Carlos's, is a cause. But this link disappears if neither*

omission is a cause, and the two differ just in that one and not the other is explanatory. [76]

The next section aims to resolve this difficulty. In summary, building on the previous section around intention, I argue that intention maters, and the moral responsibility for both intentional doing and intentional allowing can be understood by employing both a causal explanation (as already discussed in Chapter 3) and a *teleological* explanation. I shall also show how the *telos* of the agent is linked to the overall rightness and wrongness of the agent's actions.

5.7. Teleological explanation

If causal explanation (as understood by Beebee) does not give us a complete picture of the ascription of moral responsibility for allowing harm, what about forms of explanation other than causal explanation? I showed in the previous chapter that there are more forms of explanation than just causal explanation (see for example, Reutlinger and Saatsi 2018)[77].

One promising way to think about non-causal forms of explanation is to make use of an Aristotelian approach to explanation, as discussed in Chapter 4 of this thesis. Aristotle actually spoke of *four* forms of explanation (Aristotle Physics II:3 - material, formal, agent-patient[78],[79] and final (or teleological) modes of explanation. I do not plan to say anything here about material or formal causes, and agent-patient causes

[76] McGrath has different character names in her story – the gardener role is played by Barry, and the Pope role is played by Carlos
[77] Though it is also true to say that most of the literature about both scientific and non-scientific explanation does refer to causal explanation.
[78] Also known as efficient explanations - though this is commoner in Humean discourse
[79] I will use the term efficient when considering Humean ideas of causation, and the term agent-patient when considering Aristotelian forms of causation.

correspond (very) approximately to our contemporary (Humean) understanding of efficient cause which I have discussed at length earlier in this thesis (though underlying differences in Aristotle's metaphysics means that there are differences in how agent-patient explanation and Humean efficient causation work).

However, I do want to argue that explanations that appeal to final causes (what we might call *teleological* explanations) have an important part to play in the grounding of the moral responsibility attributable to allowings (and as it happens, doings as well, but I will return to that topic towards the end of this chapter).

What does teleological explanation involve?

A teleological explanation of an agent's behaviour explains the behaviour by appealing to the goal or the aim of an agent (the agent's *telos*), to which the behaviour is directed. The goal explains the behaviour. Teleological explanations seem to be a common and intuitive mechanism for providing explanations and helping us to understand the world. They distinguish between things that we do on purpose, and things that just happen to us. We do often give explanations of action that are teleological in form – things we do on purpose seem often to be done for some further reason – in order to achieve some aim or goal.

Here are some examples of teleological explanation from human behaviour:

Why are you going upstairs? – *in order to get my shoes*.

Why are you walking round the park? – *in order to improve my fitness.*

Whereas, by contrast, for many things that are non-intentional there is no further aim to mention:

Why did you cough? - I don't know, it just happened.

Why did you push that man over? I didn't intend to – I just lost my balance.

A typical teleological explanation therefore involves an agent, the agent's behaviour, a teleological connective (for example, *in order to* (achieve this goal), or *because of* (the agent's aim), or *for the purpose of* (achieving a goal), and the state of affairs which constitutes achieving the goal to which the behaviour is directed.

Here is an example of a teleological explanation:

> *John (agent) walked to the fridge (behaviour) in order to (teleological connective) get a snack (state of affairs which constitutes the goal or aim, which is usually conceived of as some kind of good).*

The behaviour of John is explained, or made intelligible, by citing the aim of the agent – to obtain a snack.

This dissertation is mainly concerned with behaviours that result in harm, and it is important to ask whether teleological explanation 'works' in the explanation of these sorts of behaviours? I believe that it does.

Consider the following (bathtub) case, which has appeared intermittently throughout this dissertation (Rachels 1975).

Consider first the 'doing' part of the case:

> *Smith stands to gain a large inheritance if anything should happen to his six-year-old cousin. One evening while the child is taking his bath, Smith sneaks into the bathroom and drowns the child, and then arranges things so that it will look like an accident.*

This behaviour can be explained (= can be made understandable) in teleological terms. The behaviour (*drowning the child*) was done because Smith (*the agent*) wanted the cousin to drown, in order that (*teleological connective*) he would go on to inherit a large sum of money (*goal, which Smith sees as a good*) Unlike the causal

explanation for this behaviour (which I have discussed earlier and involved the fact that Smith caused the cousin to die) this type of explanation does not rely on a causal story; rather the explanation relies on coming to understand Smith's aims and goals – to get his hands on the inheritance, with the further goal to be rich, or at least self-sufficient. This is a typical teleological explanation with a typical teleological order, analogous to Anscombe's 'pumping man' account discussed earlier in this chapter.

The teleological order is as below:

A: Drown the child *in order to*:

B: Get the inheritance *in order to*:

C: Get rich

Does this sort of explanation work for the allowing part of the case? I argue that it does. Consider the second (allowing) part of the 'Bathtub' case:

> *Jones also stands to gain if anything should happen to his six-year-old cousin. Like Smith, Jones sneaks in planning to drown the child in his bath. However, just as he enters the bathroom Jones sees the child slip and hit his head and fall face down in the water. Jones is delighted; he stands by, ready to push the child's head back under if it is necessary, but it is not necessary. With only a little thrashing about, the child drowns all by himself, "accidentally," as Jones watches and does nothing.*

In this part of the case, I have argued, Jones (*the agent*) allowed the cousin to drown (*behaviour*) in order to (*teleological connective*) inherit the fortune (goal, which Jones sees a good to him).

Importantly, Jones *only* allowed the cousin to drown – he did not genuinely *cause* the cousin to drown. So there is no genuine causation story available to explain this

example of allowing harm (though there is a causal explanation available, of the sort that Beebee/Lewis would favour). But there is a teleological explanation available. Jones's behaviour is (teleologically) explained by understanding his aim that the cousin should die, which in turn is explained by his goal of inheriting money (and becoming rich, financially secure etc etc).

Here is the teleological order for this sequence:

A: Let the child drown *in order to*:

B: Get the inheritance *in order to*:

C: Get rich

Teleologically, there is no difference between what Smith did and what Jones allowed. The agent's aims can be achieved just as well by employing an allowing, as by employing a doing.

Some cases are more complex than this. Consider this case:

> *Jack is a A&E doctor who has just finished a very demanding shift. As he drives home, he notices someone lying on the floor, looking unwell. He could stop but he is tired and wants to get home to his tea and to sleep. He agonises for a while, but eventually chooses not to stop, and instead continues homewards where his tea is waiting.*

This case is not as clear cut as the 'Bathtub' case, as it lacks the clear goal of inheriting a large financial reward. Presumably Jack has several aims – he may aim at the well-being of the patient (perhaps his medical training has kicked in, together with some moral sensibilities), and perhaps he also aims at avoiding criticism or even official censure. But the *dominant* aim seems to be to get home to eat and rest; and this dominant aim has the resources to explain (teleologically) his failure to

stop and help. By dominant aim, I mean that this appears to be the most plausible way of explaining the action that Jack actually performs.

I want to stress that in both the 'Bathtub' case and the A&E doctor case, the aim explains (teleologically) the actions of the agents in the story. This is a different explanation than the causal explanation of the actions, but still important.

5.8. Does teleological explanation reduce to causal explanation?

One challenge to the claim that teleological explanation is an adequate explanation for allowings (or for doings as well, for that matter) is the worry that teleological explanation simply reduces to causal explanation. As Sehon (2013, p122) says:

> *The true controversy concerns whether action explanation is irreducibly teleological. Many philosophers of mind would claim that talk of goal-direction and purpose is a useful heuristic but is ultimately to be cashed out in other terms.*

This hypothesis has the attraction of simplicity – rather than there being two explanations (causal and teleological) of a behaviour, we can manage with just one explanation (a causal explanation). Donald Davidson, for example, (1963 p3), argued that reason explanations (including teleological explanations) can best be seen as a causal explanation where the goal of the agent is caused by one or more mental states. For example:

> *John went to get a snack because he had a desire (a mental state) for a snack, and he has the belief (another mental state) that snacks were in the kitchen.*

Or, to use the language of intention:

John went to get a snack because he had the intention (= some kind of mental state, or perhaps the strongest mental state) to get a snack, and he knew that snacks were in the kitchen.

Under this account, aims and goals drop out of the explanation. Intentional actions are simply *caused* by an appropriate mental state or combination of mental states. The account of causal mechanisms (desire/belief pairs, or intentions understood as mental states) has nothing to do with the telos of the agent. To say that the agent has a telos is simply shorthand for the claim that the agent has an appropriate mental state and that this mental state *caused* the particular behaviour. Under this account, an agent A did B in order to C iff A had an intention to C and this intention *caused* her to B.

Why might this not be an acceptable story?

One possible reason for this is the difficulty we have had so far in understanding the moral responsibility of allowings – we have seen that a causal explanation offers a very attenuated explanation for harms that related to allowings, and this attenuated explanation turns out to be unable to ground moral responsibility for these harms. Furthermore, reducing the teleological story to a heuristic, offers a very thin explanation of human agency. Goal-direction seems to matter for human action; human beings make plans and reason teleologically. Eliminating teleological explanation reduces our understanding of the action. Indeed the explanation of intentional action might be the 'other way round' from the explanation offered by Davidson. It may be that agents have desires and intentions purely because they first have goals and aims.

Another obstacle to the idea that teleological explanations can be reduced to causal explanation is the challenge posed by *deviant causal chains*. There are several

examples of deviant causal chains in the literature, but reflection on the genesis of actions suggests that such cases are myriad in everyday life. I cite four widely quoted examples below.

Example 1 (Davidson 1963, p16)

> *A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and want might so unnerve him as to cause him to loosen his hold, and yet it might be the case that he never chose to loosen his hold, nor did he do it intentionally.*

Example 2 (Adams and Mele (1989 p519)

> *A philosopher intends to knock over his glass of water in order to distract his commentator. However, his intention so upsets him that his hand shakes uncontrollably, striking the glass and knocking to the floor.*

Example 3 (Davidson 1963, p16)

> *A man may try to kill someone by shooting at him. Suppose the killer misses his victim by a mile, but the shot stampedes a herd of wild pigs that trample the intended victim to death. Do we want to say the man killed his victim intentionally?*

Example 4 (Chisholm 1966, p28)

> *A man whose intention to kill his cousin made him so agitated that he drove too fast and accidentally killed a man who turned out to be his cousin.*

What do these examples have in common? They are all cases where the intention caused the outcome, but in all these cases we would be very reluctant to characterise the actions performed as *intentional* actions. Take the climber case

(Case 1). The climber wants his colleague to fall to his death. He believes that if the rope is cut then the colleague will fall to his death. The combination of the desire and belief (or the mental state of intention) causes the climber's body to move in a particular way such that the intended state is achieved (the second climber falls to his death). But it does not seem that this is enough to make what the climbers did an intentional action – seems more like and accident than an intentional action. The causal story belief/desire story does not seem to be sufficient for an action to be called intentional.

The same is true of the philosopher case, the shooting case and the driving case. Something is missing from the account of intentional action – causation is not enough, and the action can only be called intentional if the action includes aims and goals.

There is an extensive literature on causal deviance (see for example, Sehon, 2013, 2016, Chapter 6) and there are many spirited attempts to resurrect the causal story in preference to the teleological story. I will not discuss these attempts here. Suffice to say that despite these efforts, the phenomenon of causal deviance continues to create problems for the causal theory of action. This worry, together with the arguments advanced earlier in this chapter seem to suggest that teleological theories of action do not reduce to causal ones.

5.9. Teleological explanation and moral responsibility?

I have argued that teleological explanation (for both doing and allowing) explains behaviour in terms of the agent's aims and goals. I have also argued, (following Anscombe) that one cannot understand what action a behaviour constitutes (whether

a doing or an allowing) until one knows the intention with which it is done, and thus the *telos* which the agent is aiming at. Implicit in this argument (as discussed earlier in this chapter) is the idea that doing and allowings are not just act-types – they are instead act-conducts (see Audi 2016) which include the intention with which the action is done.

Furthermore (again following Anscombe) intentional doings and allowings are characterised by a teleological sequence. An agent does A in order to do B and does B in order to do C – and so on. Teleological sequences can be long or short, and can even consist of just one item – the answer to the 'Why?" question might immediately cite a final good. Whether long or short, the teleological sequence 'runs out' – so that there is no further reason to be offered to explain the agent's behaviour – there are no more 'Why?' questions to answer. Note that the 'Why?' question must always be answered by something good about the object of the intention – something good for the agent (though not necessarily morally good). But understanding the teleological sequence (which has been explained by the teleological explanation) allows one to fully understand the intentional behaviour – be that a doing or an allowing.

5.10 From teleological explanation to moral responsibility

How does the aim or goal of a teleological sequence help us ascribe moral responsibility for the doing or allowing>

I suggest that in the example given above, the drowning of the cousin gains its negative moral valence from its linkage with the (evil) aim of ending the life of the child for financial gain. The same linkage makes the cousin blameworthy – whether

he has done something (drowned the child) or allowed the child to drown. Note that the linkage is the same whether the cousin achieves his financial gain via a doing or an allowing. Hence both cousins are blameworthy even though they have achieved their goal in different ways – one via a doing and one via an allowing.

What makes the agent blameworthy in both cases is that the act is carried out with a *telos* (because the act is intentional), and it is that telos which is the morally accessible matter – that makes the blameworthy or praiseworthy. (In Aristotle's terms it is the *arche* of the action).

In both cases, the cousin who has this intention has the view that the action is good because it is a means to attaining wealth, which he mistakenly means is an ultimate good. This is a mistaken belief because it leads him to believe that killing an innocent child can be good because of its instrumental role in acquiring the fortune – when it is obviously bad in itself.

It is worth noting that harm that is produced accidentally is not part of the teleological sequence – the behaviour is not linked to the telos in the same way as it would be if the behaviour was intentional. If the cousin had no intention to harm the child, but slipped and banged his head on the side of the bath, thereby losing consciousness briefly, the harm to the cousin might still occur, but the cousin would not carry the same) moral responsibility because the cousin did not have the telos of killing the child.

Thomas Wedgwood sees the teleological sequence in slightly different way to Anscombe's idea of a teleological sequence. Wedgwood (2011 p 20) expresses this in terms of agential involvement – there is a greater degree of agential involvement in bringing about a harmful state of affairs if one has a goal of bringing about this harmful state of affairs (compared to a harmful outcome that occurs accidentally).

Plausibly, the moral responsibility for that behaviour tracks the agential involvement in that behaviour. Harms that arise by accident, although still harms, do not carry the same moral responsibility compared to harms that are aimed at.

Let's summarise the argument so far. It seems that teleological explanations exist and exist in an irreducible way (They cannot simply be reduced to causal explanation). They are commonly used and are required in explaining intentional actions. They provide a distinctive answer to a 'Why?' question with respect to such actions - an explanation which makes the act rationally explicable. Teleological explanations apply for both doings and allowings. They complement but do not replace causal explanation for doing and allowing harm. Doings are morally assessable because of their role in genuine causation and their role in teleological explanation. Allowings are morally assessable because of their role in causal explanation and their role in teleological explanation.

There is thus a difference between the grounding of moral responsibility for doing and allowing harm. This is consistent with the intuitive impression that there does seem to be a difference in moral responsibility between the two behaviours, and the difference between teleological explanation plus causal explanation (for allowing harm) and teleological explanation plus genuine causation (for doing harm) can explain this difference in moral responsibility.

This is promising, but there is much that is not yet explained. There needs to be clarity about how teleological explanation and causal explanation interact, and it is to this I will now turn.

5.11 More about telos

I said earlier in this chapter that the rightness or wrongness of a conduct is partly dependent on the *telos* or aim of the agent. So in the grandmother case discussed in an earlier section, the *telos* of Jack was a good *telos* (to cheer up his grandmother) whereas the *telos* of Jill was a bad *telos* (to get rich by being included in the will). The difference in *telos* explains (by means of a teleological explanation) the difference in the rightness or wrongness (or blameworthiness) of what Jack and Jill did – visit their grandmother. We can link the wrongness of a particular action to the *telos* with which the action is done (the telos at which the action aims).

Whether an agent's *telos* is good or bad is just one way of assessing *telos*.

Another way of thinking about *telos* (which will be helpful in understanding some of the practical cases later in this chapter) is to see it in terms of *role*. The fact that an agent has a particular role may generate a *telos* for that agent which other agents (who do not have that role) do not possess. For example, the *telos* of a doctor is to save lives, reduce suffering etc. If a doctor fails to do these things, then they will be blameworthy – even if a layperson who also failed to do these things would not be morally responsible in the same way (they may have some kind of responsibility, derived from their *telos* as a good citizen). Similarly, a lifeguard has a *telos* of rescuing struggling swimmers, a vet has a *telos* of caring for sick animals – and so on.

There thus seems to be at least two ways of assessing actions by means of the agent's *telos*. One way is in terms of good *telos*/bad *telos*. A good *telos* (all things being equal) will tend to make actions that aim at that good *telos* also good. For example, the good *telos* of '*cheering up your grandmother'* makes the action of '*visiting your grandmothe*r' *also good.* In the same way, the agent who visits their

grandmother in order to get cheer her up is praiseworthy. By contrast, the bad *telos* of *'getting money in your grandmother's will*' makes the same action of '*visiting your grandmother'* also bad. In the same way, the agent who visits their grandmother in order to get rich is blameworthy.

The second way that telos impacts on the moral assessment of actions, is via the role of the agent. An agent who has a particular role (father, mother, citizen, doctor, lifeguard etc etc) is blameworthy if his *telos* is not congruent with his particular role. A doctor who fails to stop and help (say) the victim of a road traffic accident is blameworthy (does something wrong) whereas an ordinary member of the public would not be blameworthy in the same situation – or at least not as blameworthy as the doctor.

5.12 Moral responsibility for doing and allowing harm: coarse distinctions and fine distinctions

As suggested above, the contrast between genuine causation and causal explanation seems to be able explain (in a very broad-brush way) the moral difference between doing and allowing harm. The wrongness of doing harm is explained by its role in genuinely causing harm, whereas allowing harm provides only a (counterfactual) causal explanation of the harm. I've argued that this explanatory difference is enough on which to base a rough and ready moral difference. An agent who intentionally causes harm is (usually) more blameworthy than an agent who intentionally allows harm, and is this way we can differentiate in very general moral terms between doing and allowing harm.

Let's call this difference a *coarse* moral distinction – it allows us to make general comments about the rightness or wrongness of doing and allowing harm, but the distinction is unable to deal with more subtle distinctions, and does not allow us to

make any comments about individual doing and allowing of harm – other than, of course, to say that ceteris paribus, doing harm is morally worse than allowing harm. But this coarse distinction is consistent with the generally accepted acknowledgment of the role of causation in the attribution of moral responsibility found in the literature on doing and allowing harm (see for example, Driver (2008) and Sartorio (2007, 2009). It is also consistent with the views expressed by Beebee (2004) and Varzi (2007) about the causal explanation story for allowings.

How then does intention fit into the picture? First of all, I want to argue that the wrongness of a doing or an allowing should be explained in both a causal way (genuine causation or causal explanation) but in a teleological way as well. So an agent would be blameworthy for a doing or an allowing not just because they caused or (in the case of an allowing) they causally explained the harm but also because they teleologically explained the harm. That is, the harm was part of a teleological sequence that led to a bad or evil telos, or was inappropriate for the role occupied by a particular agent (as described for a doctor or lifeguard above).

What I would also like to suggest is that intention acts to allow the agent to make fine distinctions and to differentiate behaviours within the category of doing harm or allowing harm. Here are some examples, which illustrate what I mean by fine distinctions.

5.13 The gardener, the Pope and the death of the plants.

Recall the problem from Chapter 3, where the genuine causation/causal explanation difference was unable to differentiate morally the behaviour of the gardener and the

Pope in letting the plants die. The problem arose because the gardener who intuitively was blameworthy for the death of the plants had exactly the same counterfactual relationship to the plants' demise as the Pope – who obviously is not in any way blameworthy for their death.

A teleological explanation can resolve this dilemma. The *telos* of the gardener is to look after plants, particular these plants, which belong to his employer. Hence he is blameworthy for their death – because it is his job to look after the plants, and the well-being of the plants is the *telos* that he should have. By contrast, the Pope's *telos* is nothing to do with plants – his *telos* is about caring for nurturing souls, and he would be blameworthy if he failed to do this.

To summarise, both the gardener and the Pope failure to water causally explains the death of the plants – the difference between the moral responsibility of the two agents is entirely derived from a non-causal, teleological explanation of the plants death.

5.14 Jack and Jill and the blameworthiness of visiting grandma.

Both Jack and Jill visit their grandmother, but we judge Jack to be praiseworthy and Jill to be blameworthy. This cannot be explained in causal terms – as Rachels argues (1986) they both did the same thing and caused the same outcomes (presumably, a happy Grandmother). But there is a huge difference in *telos*, and different teleological explanations operating in the Jack case and the Jill case. Jill's *telos* is bad – her aim is to get rich, and not to genuinely care for her grandmother. Jack, by contrast, has a good *telos* – he aims at cheering up his grandmother. So two actions, with the same causal explanation, but very different teleological

explanations. And it is the teleological explanation that determines the rightness or wrongness of the action.

5.15 The absence in some cases of a moral difference between doing and allowing harm

Usually there is a moral difference between intentionally doing and allowing harm – the doing of harm is more blameworthy than the allowing of harm. However, sometimes, this difference is absent. For example, in the 'Bathtub' case, we confidently conclude that there is no difference between the conduct of Smith (who kills the child) and Jones (who merely allows the child to die) – they are both equally abhorrent. How can this be explained? After all, the genuine causation/causal explanation suggests the opposite – a moral difference should be present, based on the genuine causation/causal explanation difference. Smith actually caused the child to die; Jones' behaviour is part of the causal explanation of the child's death, but he did not cause the death of the child. This seems anomalous.

I suggest that a teleological explanation might be helpful here. Both Smith and Jones have the same heinous intention with a truly awful *telos* – they are killing a human being purely for monetary gain. So the teleological explanation of the conduct of Smith and Jones suggest that they were both extremely blameworthy. This 'swamps' the much smaller difference in blameworthiness secondary to their differing roles in causal explanation. Their *telos* is so wrong that a moral difference between the two behaviours cannot be detected.

Interestingly, Trammell's comment ((1976, p132) on Michael Tooley's (1980, p104) case is not dissimilar from my explanation. (Tooley's case is related to the moral

difference between blowing someone up (a doing) and failing to warn about the presence of a bomb (an allowing)) is not dissimilar from my explanation.

Trammell suggests:

> *Jones's extreme hatred for Smith and his cynical joy at seeing Smith blown to bits have a "masking" or "sledgehammer" effect, which makes it difficult to evaluate the significance of the distinction between negative and positive duties. The fact that one cannot distinguish the taste of two wines when both are mixed with green persimmon juice does not imply that there is no distinction between the wines.*

This captures the idea that if an intention is bad enough, then the causal difference simply drops out of the moral mathematics.

5.16 A note about responsibility for unintentional allowings

So far in this thesis, I have deliberately considered only intentional allowings – allowings that are part of the teleological chain between the agent's action and their aim. But many allowings, including many allowings that lead to harm, are not like this. Many allowings are unintentional. Furthermore many (but not all), unintentional allowings are considered blameworthy. How can this blameworthiness be attributed to unintentional allowings?

I have introduced a causal explanation story for attributing moral responsibility to intentional allowings. I have also proposed a teleological mechanism for attributing moral responsibility to intentional allowings. But for *unintentional* allowings, neither the causal nor the teleological strategy seems appropriate. Another concern is the need to distribute blame to unintentional allowings – some allowings seem to be

blameworthy, and some appear not to be blameworthy. How does one differentiate between these two classes of allowings?

Let me present an example.

> A junior doctor (Jack) attends to a patient in the Accident and Emergency Department. The doctor is quite tired after a week of night shifts, and is looking forward to his holiday abroad, which will start immediately after this consultation. Both he and his partner have been planning for this holiday and Jack is already thinking about how warm the sun will be.
>
> The patient does not seem to the doctor to be too unwell, so he sends the patient home, with a provisional diagnosis of a viral infection. Unfortunately, the patient is developing a purpuric rash, characteristic of meningitis, which Jack fails to spot. Over the next 12 hours the patient deteriorates and has to attend the hospital again, this time with a diagnosis of meningococcal meningitis.

This case is an example of an unintentional omission, and unintentional allowing of harm. The deterioration could have been prevented but was not. But Jack certainly did not want the child to become more unwell, and so this is clearly not an intentional harm. Most commentators would want to attribute blame to Jack, perhaps because he was not careful enough in his assessment.

There is a problem here for my earlier ‘linkage’ account of the teleological grounding of harm related to allowings. Recall that I argued earlier that there is a teleological ‘linkage’ between an intentional allowing and the telos that the agent is aiming at. If the telos is a ‘bad one, then this makes the allowing blameworthy (as in the ‘Bathtub’ case). But in this case, there is no intentional behaviour, no ‘bad’ telos that the agent

is aiming for. Jack certainly does not want the patient to become unwell. This case seems very different to the 'Bathtub' case.

I want to argue that the telos in this case (mentally preparing for a holiday) is not in itself a 'bad' aim for this doctor – nor is the aim of pleasing his partner (they are actually both very laudable aims). What is problematic, however, is that the aim of preparing for a holiday (and pleasing one's partner) is the *wrong telos* for this particular situation. He *should* have had the aim of carefully assessing the patient (the 'right' telos), but his eagerness to plan for his holiday and to please his partner lead him to pursue the wrong goal.

So the teleological sequence account still has purchase here; we just need to be clear about what are acceptable aims for the agent. The doctor in this story is blameworthy because he has the 'wrong' *telos*. He is not aiming at a harmful *telos* as such (as in the Bathtub case) but his intentional pursuit of relaxation and approbation leads to a harmful outcome.

I suggest that an account of whether there is a 'wrong' *telos* (or not) might provide a theoretical basis for whether it is appropriate to attribute blame to agents for unintentional allowings of harm. Had Jack had the 'right' *telos* whilst assessing the patient, and had based his actions on that *telos* (performing an assessment that would be likely to achieve that *telos*) then it would be difficult to attribute blame to him for the way he dealt with the patient. The 'right' *telos* is the one that is appropriate to his role – which is the role of an emergency doctor caring for patients who are potentially critically unwell. There are goals that a doctor needs to be aiming at in order to perform his role properly – these *tele* are part of the definition of what makes a good doctor. If a doctor fails to aim at these goals, then this means he acts in an unprofessional way, and he or she may even represent blameworthy

negligence. By contrast a really good doctor is always cognisant of the correct *telos*, focussed on these goals, striving to achieve them. In the middle of the spectrum, an average doctor will be less focussed but perhaps without being blameworthy. The negligent doctor's aims will not be the same as the fundamental aims of the profession. Indeed, a negligent doctor reminds one of the indifferent gardener in Anscombe's story (1963, p 41/42):

> *I didn't care about that, I wanted my pay and just did my usual job.....I didn't care tuppence one way or the other for the fact that someone has poisoned the water, I just wanted to earn my pay without trouble by doing my usual job – I go with the house, see? And it doesn't matter to me who's in it.*

5.17 In conclusion: the difference between doing and allowing harm.

In an early part of this dissertation, I presented a challenge; what might explain the moral difference between doing and allowing harm? I have reached a tentative conclusion; moral responsibility for doing harm is grounded in both genuine causation and in teleological explanation, whereas moral responsibility for allowing harm is grounded in causal explanation and teleological explanations.

I also posed the question of how to explain the blameworthiness of the gardener but not the Pope for the death of the plants. This too can be explained teleologically; the gardener's telos (as he had responsibility for the well-being of the plants as part of his role as a gardener) was to keep the plants well. The Pope's telos, by contrast, is about saving souls not plants, and it was no part of the Pope's telos to water those plants. This explains the difference between gardener and Pope, and it goes some way to resolve the profligate ontology problem for allowing harm. The problem is not related to causation but is related to the teleological explanations of allowing.

If the existence of an intention under a description, motivating what is done, (or sometimes what is allowed to happen) is what makes a piece of behaviour constitute an action, then any intentional behaviour, whether it is (what we have been calling so far) a '*doing'* or (what we have been calling so far) an '*allowing'* would qualify as an action. And if this the case, then this would explain why (to take one example) there is no moral difference between the behaviour of Smith and Jones in the 'Bathtub' case. In Anscombe's terms, what Smith and Jones achieved (the death of the child) would be intentional in both cases (under the description; '*ending the life of the child'*), and therefore both the doing and the allowing would be, in the relevant sense, intentional actions. Whereas, for (say) in the Foot Rescue cases, one behaviour (the case when the rescuers ran over the trapped individual) would be intentional under the description: '*killing the trapped individual'.* The other case would be intentional under the description of '*hurrying to the rescue scene'*, but not for the description '*killing the trapped individual'*. This motivates the moral difference between the two behaviours and hence provides an explanation of the moral difference between the harm produced in Rescue 1 and Rescue 2.
Importantly, both the doing action of Smith and the allowing action of Jones are intentional actions – but they differ in the structure – as discussed in an earlier chapter.

CHAPTER SIX - THE MORAL DIFFERENCE BETWEEN DOING AND ALLOWING HARM – SOME PRACTICAL (MEDICAL) IMPLICATIONS

6.1 Introduction

At the beginning of this thesis, I introduced my main claim - that there was a moral difference between intentionally doing harm and intentionally allowing harm. But this difference was a *variable* difference - sometimes it was present, and sometimes it was absent. This claim contrasts with other claims about doing and allowing harm in the literature. For example, some philosophers of a broadly consequentialist bent assert that there is *no* moral difference between doing and allowing harm (for example, Singer 1993, Rachels 1975). Other philosophers argue that there is *often* a moral difference between doing and allowing harm (for example Kamm 2007, Woollard 2015).

Making this '*variable difference'* claim faces one immediate difficulty. This difficulty was identified by Norcross (2003, p 451). Norcross observed that it was very easy, by commenting on pairs of examples, to claim that sometimes a distinction is morally relevant and sometimes not. Norcross went on to say that:

This [claim] has no force. It must be accompanied by an explanation of why the distinction is morally relevant in some cases, and not in others. This is much harder to provide (2003, p 453)

To provide such an explanation is the challenge that I have accepted in this thesis. I have developed an account of doing and allowing harm that explains why sometimes there is a moral difference between doing and allowing harm, but sometimes this moral difference is attenuated or even absent. To do this, I have argued that the moral difference between doing and allowing harm can be explained by employing two different sorts of explanations – a causal explanation and a teleological explanation.

In this chapter I apply my account of this difference to certain practical ethical problems from the area of medical ethics. The reason for presenting this application is twofold. Firstly, I want to show that my theory can contribute to the understanding of several problematic (but not uncommon) practical medical ethical dilemmas which feature extensively in the medical ethics literature (and which I have encountered in my own medical practice). I want to show that my account is not just of theoretical interest but has practical explanatory value. Secondly, and equally importantly, I want to show that the conclusions reached by application of my account cohere with plausible pre-theoretical positions within the medical ethics literature. This is obviously not the final word in the matter; after all, pre-theoretical views are not always correct. But it would be a serious problem with any account if there was no coherence at all between pre-theoretical views and the output of a theoretical account. Conversely, it would be encouraging (though not definitive) if the conclusions produced by application of the theory showed a strong coherence with pre-theoretical view about practical ethical scenarios.

In this chapter, I want to set out my main claims about a range of practical cases, which all contain issues about intentional doing and allowing harm.

These are:

- End of life care, including the moral difference between euthanasia and withholding life sustaining treatment
- Termination of pregnancy, including the moral difference between killing the foetus in order to save the life of the mother
- Honesty in medical consultations, including the moral difference between lying and allowing the patient themselves to reach a false conclusion

6.2 Ethical issues in end-of-life care

In this section, I plan to cover the following issues, all related to doing and allow harm.

- The nature of euthanasia. Is euthanasia always a doing?
- The nature of withholding and withdrawing life-sustaining medical treatment (LST). Are withholding and withdrawing life-sustaining medical treatments allowings? Are withholding and withdrawing life sustaining medical treatments always allowings of *harm*?
- The causal question. Does withholding/withdrawing life sustaining treatment always *cause* the death of the patient?
- The moral difference question. Is there a moral difference between euthanasia and withholding/withdrawing life-sustaining medical treatment?

I will consider these questions in the following sections.

6.2.1 Euthanasia can be a doing or an allowing

The usual mental picture of euthanasia is of a doctor ending the life of a patient in great pain or suffering by means of the administration of a lethal injection.

So, for example, in the famous Cox case (R v Cox, 1992), Dr Cox injected Lillian Boyes with Potassium Chloride in order to end her life and end her suffering from advanced Rheumatoid Arthritis. Described in this way, it seems that euthanasia is a paradigmatic example of an intentional doing – in terms of my account it involves genuine causation and intention. Dr Cox *caused* the patient to die (but giving her a lethal injection) and intended the patient to die (because that was his aim or telos). But this is not always the case. I argue, based on my account of the doing/allowing distinction, that euthanasia can also be an allowing[80]. In my account of the distinction between doing and allowing, allowings do not involve genuine causation, but they do involve causal explanation and intention. So a doctor who *could* provide a treatment that would delay death but intentionally fails to do so in order to allow the patient to die (for the patient's good) would also be committing euthanasia. Such a process might be called passive euthanasia, to contrast it with active euthanasia such as that committed by Dr Cox. Because euthanasia can be both active and passive, it seems that what is important in euthanasia is not the genuine causation of death (which is only present in active euthanasia) but the *intention* that the patient should die.

Despite this, when we speak of euthanasia, we often (wrongly) think only of what is characterised above as *active* euthanasia. Why is this the case? I believe that this confusion occurs because withdrawal and withholding treatment occurs for a wider range of reasons other than those that apply in cases of passive euthanasia. For example, many treatments are futile and do not benefit the patient; withdrawing and withholding such treatments would not delay death, nor would they benefit the

[80] Most definitions of euthanasia allow for the death of the patient by an allowing as well as by a doing. So, for example, Wilkinson, Herring, Savulescu (2020 p 229) offer – 'euthanasia – to intentionally cause *or permit* the death of someone, for their sake'

patient in any way. Similarly, many treatments are burdensome - so the intention behind withdrawing and withholding such treatments would be to remove this burden, but not to prevent or delay death. Or the patient may have refused the particular life-sustaining treatment – this can occur because the *patient* intends his own death, but it does not follow that the doctor or health care professional has the same intention (the doctor may simply aim at respecting the patient's autonomous wishes, and might be disappointed if the patient does in fact die). The above three examples show that it is possible to withdraw and withhold treatment without the intention of ending the patient's life[81].

My account also provides an explanation for the difference between euthanasia and murder – euthanasia involves a *telos* that is at least *potentially* good – ie the relief of suffering for the patient[82]. Murder shares the causal and teleological *structure* seen in euthanasia (the murderer both intends and causes the death of the victim), but there is no redeeming positive *telos* in murder – the *telos* is unredeemingly evil.

6.2.2 Withholding and withdrawing life-sustaining treatment (LST) counts as an allowing of harm

My account of doing and allowing harm suggests that both withholding and withdrawing LST count as *allowings* of harm, rather than doings of harm. This is because the central notion of allowing involves a failure to prevent harm. I argue that withdrawal and withholding LST do not *genuinely cause* harm, and the absence

[81] And because, as moral responsibility tracks intention, the withholding and withdrawal of treatment may not always be blameworthy (as in the examples above). It is also important to state that non-intentional withholding and withdrawing can also be blameworthy – if, for example, the withholding and withdrawing is done negligently, for example (see Chapter 5 of this thesis).

[82] Some ethicists, for example, argue that when a patient is experiencing intense, incurable suffering, such as might occur during a terminal illness, then ending the patient's life is actually a benefit to the patient rather than a harm to that patient.

of this property means that (on my account of the distinction) they are allowings rather than doings.

6.2.3 Withholding and withdrawing LST does not cause death

One important question within end-of-life care is whether withholding or withdrawing of life sustaining treatment (LST) *causes* the death of the patient. The question is important for reasons given in the previous section; on my account of doing and allowing, whether the action *causes* the death of the patient (rather than just causally explains the death) differentiates between a doing and an allowing, and also contributes to the ascription of moral responsibility for the action.

So firstly, what do we mean by *withholding* life sustaining treatment? Withholding LST means to intentionally fail to provide treatment that, if administered, would prolong life. One example might simply be refraining from offering chemotherapy to a patient with cancer who has already had several courses of simpler chemotherapy and does not want to continue with more treatments. The oncologist stops chemotherapy with the patient's agreement, and the patient dies shortly afterwards. With continued chemotherapy, the patient might have lived for several more months. Or, in another example, the doctor fails to offer cardio-pulmonary resuscitation CPR) to a patient with co-morbidities. With CPR, the patient may have lived for several more months.

Next, what do we mean by *withdrawing* life sustaining treatment? This involves the removal of already established treatments which, if continued, would prolong life. An example might be the withholding of fluids and nutrition in patient with a persistent vegetative state, such as the Tony Bland case (Airedale NHS Trust v Bland, 1993). In the Bland case, fluids and nutrition were supplied via a naso-gastric tube but eventually (at the direction of the court) were discontinued. Mr Bland died about ten

days later but could have continued living for perhaps another 30 years if the treatment had not been withdrawn.

Another example is the switching off a ventilator in the Re: B case (Re: B (Adult: refusal of treatment) 2002), where Ms B (a lady with quadriplegia and breathing difficulties due to a spinal arterio-venous abnormality) would have continued to live (albeit dependent on artificial ventilation) for several years had the ventilator not be discontinued.

So do these types of intervention *cause* the patient to die?

My position is that as withdrawal and withholding of LST are types of allowing (at least when they are done with the intention of ending the life of the patient, ie as passive euthanasia) then these interventions *do not* cause the death of the patient. They merely causally explain the death of the patient (had the patient not had their treatment withheld or withdrawn they would not have died).

However, some philosophers have argued that withholding or withdrawing treatment does genuinely cause the death of the patient. Garrard and Wilkinson (2005, p66) are examples of philosophers that take this line; they mount a strong argument in favour of this position. Here is how their argument progresses. Firstly, based on examples, they make a general assertion that allowings can indeed cause harm. They give the examples of a student failing to study and hence causing themselves to fail the examination, or a driver who fails to indicate right and causes an accident. They then move on to give medical examples, closer to the situations we are considering here, where (they claim) omissions can kill. A doctor fails to prescribe antibiotics to a patient with a raging fever[83] – this failure to prescribe causes the patient to die. Or a child is drowning in a shallow pond, and a passer-by who does

[83] A case of blameworthy negligence

not stop because they are rushing to a football match fails to pull them out – this failure to help (argue Garrard and Wilkinson) causes the child to drown.

My response to these arguments is simply to deny the intuitions suggested by Garrard and Wilkinson. What caused the student to fail their examination? – I would argue that she failed because she was in a state of ignorance on the day of the examination - which could have been prevented by more assiduous preparation. Her failure to prepare did not cause her to fail, but had she chosen to work harder there would have been a preventative anti-cause (knowledge!) which would have stopped her failing. What caused the patient to die of infection? – the presence of sepsis certainly, but not the failure of the doctor to prescribe, which would have prevented the death but did not cause it. What caused the child to drown? Not the lack of help from the passersby, but the presence of water in the child's airways. What seems to be happening in these examples is a confusion between causation and culpability. My account suggests that the above agents did not cause the bad outcome but were blameworthy for the bad outcomes. Why were they culpable? According to my account, they were culpable for two reasons. Firstly, because their behaviour causally explained the bad outcome, and whilst causal explanation is less culpable than genuine causation, it carries some culpability. But in addition, these agents (I suggest) have a wrong *telos*. The *telos* of the doctor should have been to save life; he should have prescribed antibiotics. The *telos* of any citizen, when presented with a drowning child, should have made at least a minimal effort to save life. The *telos* of a student should be to prepare for exams. So doctor, passerby, student all were blameworthy – not because they caused the harm (they didn't though they causally explained the harm), but because they had the wrong *telos*.

This position is supported by a separate series of arguments provided by Dan Callahan (2012). Callahan argues that there is a fundamental *difference* between withholding treatment and causing the death of the patient. So, for example, a lethal injection will kill both a well patient and a sick patient, but a failure to offer treatment will have no effect on a well person.

Furthermore, says Callahan, when treatment is withdrawn (or never started), the patient is actually killed by the underlying lethal disease which is still present. The disease, not the withdrawal or withholding of treatment, is the physical cause of death. This says Callahan can easily be shown by the results of an autopsy. Furthermore, (2012, p118) the effect of both the treatment and the withdrawal of treatment is simply to alter the *timing* of death.

> *'.... when it is argued that there is no difference between killing and letting die, a significant fact has been forgotten: that we have decided to act 'as if allowing a patient to die is a form of physical causation equivalent to directly killing a patient.*
>
> *.....what is a fact is that terminally ill patients whose lives have been medically extended live longer lives than those who have not had such treatment. They thus die at a later time than would otherwise be the case.... if not killed directly they will eventually die as a result of their temporarily arrested lethal disease either because treatment was stopped or because even the best available treatment could no longer stay the hand of death'*

Callahan compares withholding treatment with shovelling snow in a blizzard. The failure of the householder to keep shovelling does not *cause* the drive to block up – the snowstorm does that. Similarly, consider a swimmer who gets into trouble when swimming out of his depth. A passerby sees what is happening and rushes in to try

and save the swimmer. Initially he is successful but with time the would-be rescuer become exhausted and reluctantly has to let the drowning swimmer go. Has the would-be rescuer killed the swimmer? The answer must surely be no. The death of the swimmer is caused by drowning – not by the exhaustion of the rescuer.

Why then is there moral responsibility for allowing patients to die? Callahan explains this using the analogy of moral rules. He cites James Rachels's case of Smith and Jones and the baby in the bathtub mentioned earlier in this essay. Jones, says Callahan, is not the cause of the death of the child in the same way that Smith is the cause of the child's death. But Jones is just as culpable as Smith, but Jones's culpability comes from the moral duty to look after the child – Jones ought to have saved the child. Jones is not the physical cause of death of the child as is Smith, but we have a moral rule (says Callahan) to save a child – it is this breaching of a moral rule that makes Jones culpable, not the physical cause. '*our moral rule says in effect that the actual physical cause of death is irrelevant we have declared it so by virtue of our moral rule.....but the physical cause of death has not been abolished; it has been overlaid by a moral rule that declares it irrelevant*'. (2012, p121). If one fails to save a drowning insect, argues Callahan, there is no moral culpability – because there is no moral rule in our society that specifies a duty to save drowning insects. For that reason, we say that the water caused the insect to drown, not the agent who fails to save the insect. This agent does not have any moral responsibility for the death of the insect.

My approach is very consistent with Callahan's view, but I explain the moral responsibility for allowings in a different way. Firstly, I argue that withdrawing/withholding LST is an allowing of death. The physician has the possibility of preventing the death of the patient - but chooses not to do so. To use

Hugh McLachlan's wording, the physician does not cause death when he withdraws treatment but fails to employ an 'anti-cause'. (McLachlan, 2008, p636). In order to explain the moral responsibility inherent in allowing death, rather than speak about duties and rules (as Callahan does) I speak of the *telos* of the physician. This allows us to judge the rightness or wrongness of the allowing. If a physician fails to treat a treatable condition, for which a competent patient requests treatment, then this is clearly a 'wrong' *telos* – a *telos* that is not concordant with the role of the physician. Smith and Jones are equally blameworthy because they have equally blameworthy intentions – they have the same telos – which is that the child should die and that they should inherit the money.

Other cases are less clear cut. For a patient with a very short time to live, or a patient who competently refuses treatment, the *telos* of the physician may be to *not* treat. Importantly, it is the *telos* of the physician that determines the rightness or wrongness of the withdrawal or withholding.

6.2.4 Withdrawing LST is not morally different to withholding LST.

Miller and Truog (2010) accept the above arguments for *withholding* LST – they agree that *withholding* LST does not cause death, and they accept that the patient dies from their underlying condition rather than the process of withdrawing LST. However, they deny this is the case for *withdrawing* LST. The reason they take this position is that they see withdrawing as an act – for example, the practitioner needs to remove a naso-gastric tube, or disconnect an endo-tracheal tube, or switch off a ventilator. And - argue Miller and Truog, these behaviours seem to be paradigmatic examples of acts or indeed of doings. They therefore differentiate withholding LST

from withdrawing LST, and also see withdrawing LST as closer to euthanasia than to withholding LST.

As these authors say – '*whereas withholding mechanical ventilation from a patient who has refused it merely allows the patient to die, this is not true of the act of turning of a ventilator*'. (Miller and Truog (2010), p5) They actually then go on to suggest that doctors are responsible for perpetuating what they call a *moral and legal fiction* in order to explain the normal practice (and moral acceptability) of withdrawal of LST – without challenging the societal norm that it is always wrong to cause someone's death. The 'fiction' is to pretend that withdrawal of LST does not involve the doctor causing the death of the patient. (my account would see the *telos* of the doctor as to provide only non-futile treatment).

They also present a second argument. Miller and Truog ask us to consider two motorcyclists (John and Sam) who are both aged 50 and both have spinal injuries following separate RTAs. John remains dependent on a ventilator, but Sam eventually is weaned off the ventilator and can breathe spontaneously. Three years after the accident, both feel their life is not worth living and want to die. John can ask for his ventilator to be discontinued and he dies. But this option is not available to Sam. If a doctor ends Sam's life (say with a lethal injection) then that doctor would be liable for Sam's death - he would be causing Sam's death and would be killing him. Miller and Truog argue that what explains John's death is the discontinuation of the ventilator, not the underlying condition. They then make the move from explanation to causation and argue that this causal element undermines the claim that the doctors allowed John to die. It seems to Miller and Truog, on the basis of comparison with what happened to Sam, that the doctors actually killed John (and did not 'just' allow him to die).

Miller and Troeg bolster their argument that the doctors have killed John, by saying that the stopping of the ventilator causes death quickly, and that John would continue living for many years (they say 10 or more) if the ventilator was not discontinued. This they feel implies that the doctors have caused John's death. Furthermore, they appeal to an account of doing that involves initiation of a sequence – the doctors who discontinue the ventilation – '*initiate the fatal sequence, as distinct from merely permitting it to continue without intervention to stop it'* (Miller, Truog and Brock, p456). Underlying their account is the thought that switching off the ventilator is an act – no different from switching off a car ignition or a light switch in one's house.

I disagree with the arguments presented by Truog, Miller and Brock. Firstly, I have already described the problems implicit in thinking of this case in terms of acts versus omissions - and have made a case that the crucial distinction is better characterised as the difference between doing and allowing. This approach is particularly salient in cases of stopping ventilation and in cases of withdrawing fluids and nutrition, and was relied on in the opinions offered by several of the Law Lords in the Tony Bland case (Airedale NHS Trust v Bland, 1993). What matters morally when considering withdrawing fluids and nutrition from *Bland*, is not the act of taking out the naso-gastric tube, but the fact that a treatment that would prevent Tony Bland from dying is not being offered. Similarly, in the Re: B case mentioned earlier, the patient's death from spinal cord damage is not being prevented when ventilation is discontinued. Thus it seems that it is not appropriate to describe the 'switching off' process of the ventilator as an 'act'. More important is the fact that discontinuing the ventilation process (or the nutrition/hydration process) is failing to prevent the death of the patient – ie it is an allowing rather than a doing. As it is an allowing rather

than a doing, the fact that there is an 'act' involved is irrelevant – some allowings involve acts and some involve omissions. The fact that it is an allowing means that death is not caused when the ventilator is switched off.

There is another argument against my position. This problem was first identified by Brock (1985, p864), and might be called the 'greedy relative problem'. Here is how Brock describes it in his paper:

> *Consider the case of a disabled but competent patient, able to survive for months, perhaps years, on a respirator, but who comes to a firm conclusion that he wants the respirator stopped and to die. The physician accedes to his request and turns off the respirator. [....] it is commonly said, by physicians and the general public, that this is merely a case of allowing the patient to die. Imagine that, before the physician could act, the patient's nephew, tired of waiting for 'the old man to die' to inherit his money, comes into the room and turns off the respirator, causing the cousin to suffer respiratory distress and then death. He nephew would be charged with murder in the law, and condemned for morally for wrongful killing. A response by the nephew that he did not kill his cousin but merely allowed him to die or let nature take its course and that his respiratory illness killed him would quite rightly be treated as a piece of specious absurdity*[84].

[84] Savulescu (2013, p257) makes a similar point in the context of withdrawal of neonatal intensive care - '*in neonatal intensive care, it is not uncommon for a doctor to withdraw artificial ventilation with the consequence that a baby dies. If anyone else performed that action, including the parents of the child, it would be viewed as an act of homicide, in some cases murder. But when a doctor performs that act, it is not murder. The active withdrawal of a life prolonging medical care (an intentional act that ills, even if not necessarily with the intention to kill) is a standard part of medical practice in relation to people who experience severe disability and suffering"*
'Whether it is right or wrong might depend on whether a doctor performs it, but the nature of some physical event cannot logically depend on the identity of the person involved'

Brock's point is that whether an act is described as 'murder' (ie a killing) or 'letting the patient die' cannot depend on who performs that act. After all, what is done is exactly the same, whether the doctor or the greedy relative performs that act. Both doctor and greedy relative perform the same movements, flick the switch in the same way and stop the machine in the same way. The same physiological changes take place in the patient, and the patient dies in the same way.

At this point in the argument, there are perhaps two responses. The first response is to suggest that the doctor and the greedy relative *both* allow the patient to die. But it seems wildly counter-intuitive to describe what the greedy relative does as an *allowing* – to do so would go against moral and legal precedent to do so. They therefore (not unexpectedly) claim that both the doctor and the relative kill the patient. In other words, the doctor's (as well as the relative's) behaviour is a doing, and the doctor kills the patient.

My account of the distinction between doing and allowing avoids this dichotomy, and, I think, provides a better explanation for what is going on in this case. Recall Anscombe's account of the pumping man – one can only understand what the pumping man is doing if one knows the *telos* of the action. Moving the pump handle up and down could be described as '*killing the inhabitants of the house'* or '*making a nice clicking sound*' or (simply) '*doing one's job and earning a wage'*. On Anscombe's account, the switching off of the ventilator by the doctor (in discussion with the competent patient) is *not* the same action as the switching off the ventilator by the greedy relative. The *telos* of the doctor is (presumably) to respect the autonomy of the patient, and perhaps to remove a burdensome treatment. The *telos* of the greedy relative is to become rich. The differing aims of act of switching off the ventilator make the actions different, and in turn make the morality of the action

different. One action is a heinous crime; one action is a well-considered therapeutic decision. My account trades on these differences – because removing the ventilator is an allowing which does not involve genuine causation (see earlier) but only causal explanation, the moral responsibility for the allowing is mainly determined by the intention of the agent who is removing the ventilator.

My approach seems counterintuitive because it identifies both the doctor and the greedy relative as allowing the death of the patient – and this seems wrong, because whilst we might be prepared to accept that the doctor allowed the death of the patient, we are reluctant to accept that the greedy relative allowed the death of the patient. We want to say that the greedy relative killed the patient, or that the relative caused the patient to die. I think there is here a confusion between causation and culpability. Under my account, the greedy relative was blameworthy for the death of the patient, in a way that the doctor is not. But the moral responsibility for the death does not arise from causation, but from the bad *telos*. In causal terms, both the doctor and the greedy relative did exactly the same thing – they disconnected the ventilator and allowed the illness to take its course and cause the death of the patient. Both conducts are allowings – but they seem to be different sorts of conducts because the *telos* is so different.

6.2.5 There is a moral difference between passive euthanasia and active euthanasia

Combining the previous sections now allows us to understand the moral difference between active euthanasia and passive euthanasia performed by withdrawal or withholding of life sustaining treatment.

Withholding/withdrawing LST lies on the 'allowing' side of the doing-allowing distinction, whereas active euthanasia lies on the 'doing' side of this distinction. Active euthanasia therefore involves genuinely causing the death of the patient, where the *telos* of the act is to end the patient's life for the patient's good. Passive euthanasia lacks this genuine causation element (it only causally explains), though shares with active euthanasia the telos of ending the life of the patient for the patient's good. The presence of genuine causation in active euthanasia, and the absence of genuine causation in passive euthanasia, makes the two behaviours morally different.

I have argued earlier in this thesis that whether a behaviour actually causes a harm is an important component in ascribing blameworthiness for that behaviour. So the presence of genuine causation in active euthanasia (but not in passive euthanasia) will (if the death of the patient is considered to be a harm[85]) add to the blameworthiness of the behaviour. How much 'additional' blameworthiness is added by the causation element will vary between different situations. This is because the overall blameworthiness of a doing or allowing is also determined by the *telos* of the behaviour. Some particularly egregious behaviours will have such a bad *telos*, that the contribution of causation to the moral wrongness of the behaviour is minimal – in a sense the blameworthiness of the behaviour is entirely determined by the *telos* (when this occurs, there will be morally very little moral difference between doings and allowings).

[85] Whether euthanasia is considered to be a harm to the patient, or a benefit to the patient, depends on one's underlying stance on the right course of action towards a patient who is suffering intractable pain (or other intractable symptoms) and the only way to relieve this suffering is by ending the life of the patient. Some consider that any action that hastens death is prima facie wrong, even if it also ends suffering.

My view is thus different from a view common amongst philosophers currently. For example, James Rachels in his famous *New England Journal of Medicine* paper argued that (1975, p 80) - '*I have argued that killing is not in itself any worse than letting die...it follows that active euthanasia is not any worse than passive euthanasia*'. Why does Rachels's account and my account reach different conclusions? Rachels bases his conclusion on the outcome of the 'Bathtub' thought experiment. I would argue that in cases like 'Bathtub' the behaviour of both Smith and Jones is so egregious that the difference in blameworthiness due to genuine causation (Smith caused the child to die; Jones did not) is swamped by the wrongness of the evil *telos* of taking a life for financial gain.

Note that my account of the difference between active and passive euthanasia relies on my account of the difference between doing and allowing. My account explains why active and passive euthanasia are *different* – but it does not adjudicate on the overall rightness or wrongness of euthanasia per se. This depends (as my account indicates) on the overall goodness of the *telos* of ending a life in order to end suffering. Some ethicists and health care professional think that intending to end a life is always wrong – that the fact that the agent *aims* to end a life is such a strong wrong-making characteristic that euthanasia will inevitably be wrong. Others might argue that the goal of producing a state of affairs where the patient is spared intense suffering or is allowed to have control over their own life is a good aim. My account cannot adjudicate between these two views – all it can do is argue that there is a moral difference between the cases of active and passive euthanasia.

6.3 Late termination of Pregnancy

6.3.1 Introduction

Many (perhaps most) of the practical application of ideas about the moral difference between doing and allowing harm relate to ethical issues related to end-of-life care; passive v active euthanasia, withdrawal v withholding life sustaining treatment – and so on. I have discussed these kinds of cases in the previous section of this chapter. There are, however, other areas of medicine where the alleged moral difference between doing and allowing harm might also be relevant. One such area is the ethics of termination of pregnancy (usually late termination) - where the choice is between killing the foetus (so that the mother's life is saved) – and allowing the foetus to continue living, with the consequence that the mother dies (usually from complications of labour).

Initial consideration of this dilemma was probably introduced by Jonathan Bennett in a 1966 paper 'Whatever the Consequences' (Bennett 1966). Bennett rejects the view that it is *always* wrong to kill the foetus (simply because it is killing) and in doing so denies that killing is always worse than allowing to die. For Bennett, the *consequences* of doing and not allowing are all that matter, and there is nothing morally different between killing and letting die (or, more generally, between doing and allowing harm) when the outcome is identical[86]. Sometimes, says Bennett, it *appears* that there is a moral difference between killing and letting die – but this difference is due to a contingent association between other differences that matter morally – such as inevitability of outcome (which is often more likely with doings than allowings).

[86] This is reminiscent of the views of other consequentialists such as Rachels, Tooley, and Singer

In my account, by contrast, I claim that there *is* a moral difference between killing and allowing to die (and between doing and allowing harm).

In the next section of this chapter, I summarise Bennett's argument. After this, I show that Bennett's argument does not achieve what it claims to achieve. I then show that my general account of doing and allowing provides a plausible explanation of the moral difference between doing and allowing harm in termination of pregnancy cases (like the one that Bennett introduces).

6.3.2 Bennett's argument

Bennett introduces his paper in this way:

> *'The following kind of thing can occur*[87]*. A woman in labour will certainly die unless an operation is performed in which the head of her unborn child is crushed or dissected; whilst if it is not performed the child can be delivered alive by post-mortem Caesarean section. This presents a straight choice between the woman's life and the child's'. (p83)*

[87] This paper was written in 1966, and with 21st century obstetric care in a developed country, I doubt whether there would be such a stark contrast between the outcomes of the two options. In 2025, the mother *and* the baby could and should be saved by prompt Caesarean section or (as a worst case scenario) hysterectomy and delivery of the baby (with the mother's life being preserved). However, in developing countries, where obstetric facilities are rudimentary, craniotomy is not uncommon, and 'destructive operations are an essential part of obstetric practice and cannot be wished away' (Xess and Shrivastava 2019). So globally, Bennett's paper, and the moral dilemma he introduces, still has relevance today. But even in developed countries, this sort of dilemma still appears in other clinical situations. Consider a pregnant woman who is suffering from aggressive cancer of the uterus (Haas, 2017, p250). The obstetrician has a choice between removing the uterus of the mother, which will of course kill (cause the death of) the foetus, even if this is not intentional. As in the Bennett case, if this is not done, and the cancer is untreated, the obstetrician will allow the mother to die. Another relevant case (USCCB, 2009, directive 45) is the pregnant mother with heart failure due to cardiomyopathy (a degenerative condition of the muscles of the heart) that is exacerbated by the pregnancy. The obstetrician has the choice between killing the child (to save the mother) allowing the pregnancy to proceed – which would lead to the death of the mother from heart failure.

Bennett sees this as a choice between securing the mother's survival (by the mechanism of killing the unborn child - versus allowing the mother to die but in doing so, saving the life of the child (via urgent post-mortem delivery).

As Bennett puts it:

> *To do the operation (craniotomy) would be to kill the child, whilst to refrain from doing it would not be to kill the woman but merely to conduct oneself in such a way that – as a foreseen but unwanted consequence – the woman died. (p83)*

Bennett implies that it is permissible to operate and kill the foetus. He is dismissive of what he calls the '*conservative'* position that - *'...it is always absolutely wrong to kill an innocent human, even in such dismal circumstances as these' (p84)* He thinks that there are no *absolute* prohibitions against killing. For Bennett, it is not always the case that killing is wrong '*whatever the consequences*' (p84). Thinking that consequences do not have the final say in the matter is a mistake, says Bennett, and lines of thought that take this view '*need to be attacked'* (p86).

What has this to do with the moral difference between doing and allowing harm? Bennett starts by arguing that the moral conflict here is between two different ways of involving human death. If we kill, thinks Bennett, '*the death is part of a killing, in the other [case – ie when the mother is allowed to die] there is no killing, and a death only occurs as a consequence of what is done*' p85). Bennett claims that the conservative needs to demonstrate that the difference between these two ways of involving death matters morally; that the difference between actually bringing about the death of a human versus only allowing the death as a consequence has moral weight, and justifies the claim that operating would be wrong.

Bennett sees the key distinction is between '*what he did'* and '*the consequences of what he did'.* In other words:

> What he did when he operated – he kills the child
>
> Consequences of what he did when he did not operate – the mother died (though the obstetrician does not kill her).

How might we explicate the distinction between '*what he did'* and '*the consequences of what he did*'. Those who espouse the view (termed 'conservative' by Bennett) will be anxious to show that there is some moral difference between 'what he did' and 'the consequences of what he did'. If this difference is present, this would allow the conservative to argue that whilst it is *always* wrong to kill an innocent it may be permissible for an innocent to die as part of the consequences of what the obstetrician did – i.e. that there is a difference in blameworthiness between killing and allowing some consequences to occur.

So what is this moral difference between operating and not operating? Bennett considers and rejects several possibilities[88]. Firstly - the degree of *expectation* with which the agent holds that the consequence will actually occur. Bennett thinks that this is the same for both the options – the obstetrician certainly expects the foetus to die if the pregnancy is terminated, but also expects the mother (perhaps not certainly, but *almost* certainly) to die if the pregnancy is left to proceed, such that the foetus can eventually be delivered safely.

Secondly, Bennett considers the objective *likelihood* that the outcome would ensue given the behaviour of the agent. Again, Bennett thinks that objectively, the foetus will very likely die if the pregnancy is terminated. In the same way, Bennett thinks

[88] Which he divides into two groups – differences that may ground a moral difference but do not apply to every instance of the obstetrical example, or differences that do apply to this example but do not ground any moral difference. The first five cases cited in the body of the text are in the first group and the final two are in the second group.

that there is a very high objective likelihood that the mother will die if the foetus is not terminated.

Thirdly (and most importantly in relation to my particular account of doing and allowing harm) Bennett considers whether the upshot was the *intention* of the agent (or not). To use Bennett's words – '*whether I made the movement partly or wholly for the sake of achieving that upshot'* (p87). Bennett does not say what he thinks is the final aim in these cases, but what he thinks is certainly *not* the aim is the killing of the child. In Bennett's words - '*if the obstetrician crushes the child's head he does so not because this will lead to the child's death or because it constitutes killing the child, but because that is his only way of removing the child's body from the woman's' (p87)*. Similarly (though Bennett does not say this) in the case of not killing the foetus, the aim is presumably not to kill the mother, but to simply allow the foetus to survive. What Bennett seems to be getting at is that the aim in each of the situations is not killing, but simply performing some obstetric intervention.

Fourthly, Bennett wonders if there is an *independent moral judgement* about the moral difference between the two. In Bennett's words – '*your condemnation of me, or perhaps your approval, may be reflected in your putting on the 'action' side of the line an upshot which an indifferent onlooker would count as merely a 'consequence'* (p91). The idea here seems to be that an onlooker who independently felt that allowing the mother to die was morally wrong would place this allowing on the doing/killing side of the distinction. Bennett (rightly I think) considers that this produces the wrong conclusion for the conservative as this will tend to minimise the moral difference between killing the foetus and allowing the mother to die.

Bennett thus rejects all four of these possibilities – he thinks that they are present in *both* the situations of killing the foetus, or failing to operate (so that the mother dies).

Hence, claims Bennett, they cannot ground a moral difference between the two behaviours in a plausible way.

Fifthly, Bennett considers the possibility of *immediacy* – in the killing but not the letting die, there is an immediate connection between what the obstetrician does with his hands and the death of the child, and this immediacy is not present with the death of the mother (the death of the mother would typically occur days or even weeks later, whereas if the foetus is killed by a termination of pregnancy then this will be immediately apparent) Unlike the four previous reasons (expectation, likelihood, intention, moral approbation) which were symmetrical between killing and letting die, Bennett accepts that there is a difference in immediacy between killing and letting die. But he denies that this difference has any moral import. Bennett says – '*I cannot think of a remotely plausible answer which does not consist solely in an appeal to an authority*' (p92). And this seems right.

This leaves us with one more way to explain any putative moral difference between killing and letting die, and this is to rely on the difference between doing harm and allowing harm. In Bennett's words – '*not operating is not killing-the-woman because it is not doing anything at all but is merely refraining from doing something'* (p92). Bennett thinks that these two behaviours are indeed different but this difference (as in the case of immediacy) makes no *moral* difference between the two behaviours. Bennett's basic assertion is that the *only* difference between killing and allowing to die relates to the way that the surgeon (perhaps the whole surgical team) *moves*. Bennett gives the example of a man, Joe, who kills a calf. He says that the way of moving that expresses killing is as follows – '*of all the other ways in which Joe might have moved, relatively few satisfy the condition: if Joe had moved like that, the calf would have died*' (p94). So Joe might have eaten lunch, or read a novel, or danced

a jig, or played tennis, or done umpteen other things. None of these movements would have resulted in the death of the calf. Only a relatively few movements would have led to the death of the calf – perhaps thrusting with a knife or shooting with a gun. By contrast, allowing (the calf) to die translates to – '*of all the other ways in which Joe might have moved almost all satisfy the condition: if Joe had moved like that, the calf would have died*' (p95). If Joe might have eaten lunch, or read a novel, or danced a jig, or played tennis, or done umpteen other things – the calf would still have died. Bennett thus sees a difference between doing (the only set of movements that would have produced that upshot) and allowing (movements other than the only set which would have produced that upshot'.

So – if the obstetrician *killed* the foetus, it means that most of the movement he could do (for example - walking, dancing, eating) will *not* involve the foetus dying and only a very small range of movements would lead to the death of the foetus – perhaps injecting the mother with an abortifacient or performing a procedure on the mother's uterus. By contrast he would have *allowed* the mother to die if almost all of the movements he made (walking, dancing, eating) would surely involve the mother dying. The more movements that would lead to the victim dying, the more likely it is to be an allowing. The fewer movements that would lead to the victim dying - the more likely it is to be a killing. The final bit of Bennett's argument is to say that this is a real difference, but it is not a *moral* difference. Why (asks Bennett) should this sort of difference between other things the agent could have done matter for the morality of a particular action?

6.3.3 Responses to Bennett.

How should one respond to Bennett? I want to argue that Bennett can be criticised on three counts.

6.3.4 Problem 1 - Bennett's account of intention

Bennett seems to offer a very sparse account of intention. He argues that the intention of the obstetrician in crushing the child's head is to act thus '*not because this will lead to the child's death.....but because this is his only way of removing the child's body from the woman's*' (p88). His argument seems to be that the obstetrician did not intend to kill the child, but merely intended the performance of an obstetric manoeuvre. This seems a grossly simplistic account. Recall Anscombe's A-B-C-D account of the man pumping at the well. The workman's intention is to: move the pump handle – in order to pump the water – in order to send water to the house – in order to poison the inhabitants. Whilst the final aim of the action (to poison the inhabitants) is important, we can only understand the whole intention by understanding the (intentional) *means* to that final aim.

Similarly, the aim of the obstetrician (one might assume) is to save the mother, but as a means to do this, one vital step is to kill the child – and this is morally significant (counterfactually, if the child is *not* killed, then the mother will not be saved, at least in the clinical scenario that Bennett describes). We can only understand the intention of the obstetrician if we understand this pathway – not just his final aim, but the means by which this final aim is achieved. Bennett wants to claim that the killing of the child is *not* an aim of the obstetrician. I disagree; it may not be a final aim, but it is a means – a subsidiary aim on the pathway to the final aim. It thus undermines Bennett's claim that the obstetrician's intentions in the two situations are the same; they are clearly not. This in turn weakens his claim that there is no moral difference between killing the child and allowing the mother to die.

6.3.5 Bennett's case comparison is illicit

Though he does not make this claim explicit, Bennett wants to challenge the conservative position that it would be wrong to kill the child, but not wrong to allow the mother to die (if this was the only way to save the child). But the way that Bennett has set up his argument seems to be illicit – he is asking us to compare the morality of two very different actions – killing a *foetus* and allowing a *mother* to die. This way of setting things up obscures the moral truth underlying the issue, and makes it an invalid comparison. It is difficult to accurately assess the moral implications of this sort of comparison. Many other moral factors intrude which might alter how one assesses the rightness or wrongness of the behaviour. For example: the mother may have dependents who rely on her – the future of the child depends on the mother being alive - the mother has an established network of friends and family who care for her. All these would tend to persuade to persuade the reader (even if unconsciously) to give greater moral weight to the life of the mother over the life of the foetus. In addition, there may be a pre-established acceptance that termination of pregnancy, even late termination of pregnancy, is morally acceptable. After all, late termination of pregnancy (at least for foetal disability) is one of the accepted reasons to abort a pregnancy in the UK (Abortion Act 1968). All these factors combine to give the impression that it is more wrong to end the life of the mother rather than end the life of the foetus (and, indirectly, undermine the idea that killing is morally worse than letting die).

How could we make a better comparison? Here is a suggestion:

Twin foetuses

> *A woman is pregnant with twin foetuses – foetus A and foetus B. Because of problems with foetal circulation, only one can survive (the two foetuses are in competition for a limited blood supply). Foetus A is in more peril than Foetus B and if nothing is done, will quickly die. This will allow Foetus B to survive and to be delivered (later) by normal delivery. Foetus B is stronger than foetus A but can be killed by a simple intrauterine injection. If this is done, then this will allow foetus A to survive, who can then be delivered (later) by safe vaginal delivery.*

This seems a better comparison that the comparison between mother and foetus that Bennett sets up in his example – foetus A can be allowed to die, whereas foetus B has to be actually killed. In terms of the wrongness of ending the life of either A or B, it seems intuitive that it is *more* blameworthy (all other factors being equal) to end the life of B (by a killing) rather than to end the life of A (by an allowing). It is an invidious choice, but if one is forced to make this choice it will be more blameworthy to kill B than to allow A to die.

This comparison seems more valid than Bennett's mother-child comparison, as it minimises the difference between the two deaths that are being compared.

6.3.6 Bennett's account of the distinction between doing and allowing harm is invalid

One of the claims made in Bennett's '*Whatever the Consequences'* article is that there is a difference between killing and letting die (or between doing and allowing. harm) but that difference is of *no* moral importance. My next step therefore is to challenge Bennet's account, to open up the possibility that there is in fact a moral difference between killing and letting die. This will indirectly support my account of

the distinction and my analysis of these cases. I argue that Bennett's '*most of the ways he could have acted'* distinction does not accurately capture the difference between doing and allowing harm. This was first pointed out by Dinello (1971, p83) and these objections were developed by Quinn (1989, p 295). Quinn asks us reflect on a scenario first developed by Bennett (1981, 55-69) himself. 'Henry' is in a perilous (though slightly bizarre) situation. Here is Henry's situation as described by Bennett and Quinn.

> *Bennett imagines a situation in which if Henry does nothing, just stays where he is, dust will settle and close a tiny electric circuit which will cause something bad - for example, an explosion that will kill Bill. If Henry does nothing, he is by Bennett's criterion positively instrumental in Bill's death. If Henry's body were activating the device - if he were depressing a trigger or conducting a current - we might see his agency as positive despite his motionlessness. But Bennett doesn't assign any such role to Henry's body. Only one of Henry's physical actions, staying still, will cause the death, while indefinitely many will prevent the death.*

Clearly this gets '*most of the ways he could have acted'* difference the wrong way round – most of the ways that Henry could move will *not* lead to Bill's death, yet it seems that Henry allows Bill to die by staying still.

Conversely, in a different perilous situation, if Henry makes any move at all, he will set off a motion detector that will cause an explosion and kill Bill. If he stays still, the motion sensor will not be activated, and Bill will not die. Here it seems that most of the ways that Henry could move will kill Bill, rather than allow Bill's death, which is again opposite to the conclusion suggested by Bennett's account.

These sorts of arguments seem to undermine Bennett's overall account of the lack of moral importance inherent in the difference between doing and allowing harm, or indeed between killing and letting die. In turn, this suggests that a different account might be worth exploring.

6.3.7 My account of the difference between doing and allowing harm (or between killing and letting die).

The three problems with Bennett's analysis and account of killing and letting die point towards the desiderata for a coherent account of the moral difference between doing and allowing harm. Such an account would point to a plausible difference between doing and allowing harm (and between killing and letting die) that explains the moral difference between the two behaviours. Such an account would also emphasise the importance of intention in understanding the ethics of doing and allowing harm.

I believe my account meets these desiderata.

The obstetrician who operates and kills the foetus both *causes* the foetus to die (because of the physical intervention he performs) and *intends* the foetus to die. By contrast, the obstetrician who does not operate allows the mother to die but does not genuinely *cause* her to die – he or she provides a causal explanation of the mother's death. The genuine cause of the death is the complication of the pregnancy (whatever this is; Bennett does not say). This is an important moral difference between the two behaviours.

There is also a difference in intention, which grounds a moral difference. When the obstetrician performs the termination procedure, he intends to kill the foetus; all the surgical intervention has that aim, and if the obstetrician fails to end the life of the

foetus, then he would have failed in that aim. Of course the obstetrician has a further aim (to save the life of the mother) but the means to this aim is an intentional killing of the foetus.

If the obstetrician elects not to intervene, and allows the mother to die from the complications of pregnancy, then the intention of the allowing is to preserve the life of the foetus. There does not seem to be an intention for the mother to die; the death of the mother is predictable but not intentional. (If by some chance the mother survived, the obstetrician would be delighted and there would be no feeling that the medical team had failed in its aim. The aim was simply not to kill the foetus, and this has been achieved).

Of course, there is some moral responsibility even if the death of the foetus was not intentional. Imagine that the obstetrician performed an ill-advised and risky procedure and mother and foetus both died. There is no intention here for the foetus or mother to die (the doctor would be delighted if both lived) but the doctor bears some responsibility for this mortality, even if it was incidental.

Additionally, in my earlier case ("Twin foetuses') this account seems to explain the moral intuitions present in the case. (I have copied the case below for ease of access)

Twin foetuses

> *A woman is pregnant with twin foetuses – foetus A and foetus B. Because of problems with foetal circulation, only one can survive. Foetus A is in more peril than Foetus B and if nothing is done, will quickly die. This will allow Foetus B to survive and to be delivered (later) by normal delivery. Foetus B is stronger than foetus A but can be killed by a simple intrauterine injection. If*

this is done, then this will allow foetus A to survive, who can then be delivered (later) by safe vaginal delivery.

It is plausible to think that it would be more wrong to kill the foetus B by injection than simply to allow Foetus A to die of its condition. This is because the obstetrician is *causing* Foetus B to die in the killing case, whereas background medical conditions are causing Foetus A to die in the allowing case. In addition, the obstetrician *intends* Foetus B to die (unless Foetus B dies, Foetus A will not be helped). By contrast, there is no intention for Foetus A to die. This is shown by the fact that if (by some miracle) this foetus did survive (perhaps the initial assessment was wrong) then the obstetrician would be delighted.

My account seems to get the difference between Foetus A and Foetus B right. What about the case that Bennett describes? At first glance my account seem to get the morality of the case completely wrong. Accepted medical practice would conclude that the right thing to do would be to abort the child and save the life of the mother. But abortion of the child would both cause the death of the child and be intentional; whereas allowing the mother to die would neither involve causation nor be an intentional act. On my account, killing the child in order to prevent the death of the mother would therefore be *more* wrong than simply allowing the mother to die.

How is it best to respond to this worry? I think there are two possible responses at this point. The first response would be to simply bite the bullet and accept that the morally preferable course of action would be to let the mother die; this is the approach taken by (for example) some Catholic commentators (see USCCB 2009). A second approach would be to build on the flaws in the way that Bennett's case is set up – that the comparison between the death of a foetus/child and the death of a

mother is too difficult to make, and that other factors, of the sort I outlined above, may modify the moral arithmetic.

6.3.8 Conclusion

The aim of this discussion has been to provide a second example of the practical relevance of my analysis of doing and allowing in cases in medical ethics. Thus I have argued that in issues related both to termination of pregnancy as well as in end-of-life care, there is a moral difference between doing and allowing harm – and this difference is explained by my account of causation and intention.

6.4 Lying and Misleading

6.4.1 Introduction

So far, I have discussed the practical implications of the ethical difference between doing and allowing harm in two areas of medicine; the moral difference between various interventions in end-of-life care, and the moral difference between various interventions related to termination of pregnancy.

There is, however, at least one more area of medicine where the alleged moral difference between doing and allowing harm appears to be relevant. This area concerns the moral difference between lying and non-lying methods of deceiving patients. The relevant question is whether there is a moral difference between deceiving a patient by lying to them (which might be seen as a doing), and deceiving a patient in some other way, such as by the omission of information, or by implying (but not actually stating) a false piece of information (which might be seen as an allowing). Such 'non-lying' deceptions have received much attention recently. For

example, the Francis report (2013) emphasised the need for candour when errors occur during the course of patient care. The Tracey case (Tracey v Cambridge University Hospital NHS Foundation Trust and another (2014) emphasised the importance of honest sharing of information about resuscitation decisions with patients (or their proxies). The Montgomery case (Montgomery (Appellant) v Lanarkshire Health Board (Respondent) 2015) recommended that patients contemplating treatment should be informed about any material risk of a treatment before giving consent to the treatment. Furthermore, the information shared should include information that the particular patient would want to know. All three of these landmark cases involved patients who had been previously deceived by their medical practitioners – though they may not have been actually lied to.

In this part of the chapter, I make two claims. I claim firstly that there *is* a moral difference between deceiving by lying as opposed to deceiving by means of a non-lying method. I claim that deceiving patients is (almost always) wrong, but deceiving patients by lying is more serious than deceiving patients by other methods. I claim secondly that the best explanation of this moral difference comes from an understanding of the moral difference between doing and allowing, as discussed in previous chapters of this thesis, and in previous sections of this chapter.
The section is structured as follows. In section (2) I will clarify the difference between lying and other forms of deception. Then, in section (3) I discuss accounts which deny there is any moral difference between lying and other forms of deceiving. In section (4) I discuss accounts that assert that there are moral differences between lying and other forms of deceiving. Finally, in section (5), I outline my account of the

moral difference between lying and other forms of deception, and show that my account is plausible and offers a satisfying account of this moral difference.

My conclusion emphasises that the moral difference between lying and non-lying forms of deception is of the same kind as the moral difference between doing and allowing harm.

6.4.2 Lying and other forms of deception; some definitions

(i) Lying

Lying is commonly understood (for example, Adler 2018, Carson, 2006) '*as asserting or stating what you believe to be false, intending to deceive the hearer into believing it'.*

This definition thus has four components:

(i) an *assertion* condition (the agent has to make a statement, either as a speech act or as a written act)

(ii) an *untruthfulness* condition (the agent has to believe that the statement they are offering is false, i.e. that they are making an untruthful statement),

(iii) an *addressee* condition (the agent has to make the statement to another person)

(iv) an *intention to deceive* condition (the agent has to have the intention to deceive).

The assertion condition precludes the use of deceiving by silence – if there is no assertion made then there is no lie.

The untruthfulness condition means that in order to lie, the agent has to believe that the statement is false. If an agent makes a statement that he believes to be true, but is turns out to be actually false, then this is *not* lying (rather, this is accidentally giving

the wrong information). Conversely, if an agent makes a statement that he thinks is untruthful (but is actually true), then this is still lying.

The addressee condition means that It is not possible to lie to someone who cannot understand your statement. It is not possible, for example, to lie to a tree, or a goldfish, for example.

The intention condition means that in order to lie one must have an intention to deceive – so making a sarcastic statement, or using irony, or acting in a play, or telling a 'bare-faced' lie - all lack the intention to deceive, and are therefore not lying (even if they are false statements). This is because in these types of communicative acts, there is no intention to deceive the audience.

(ii) Intentional (non-lying) deception

Lying is not the only way to deceive. Whereas lying is done by intentionally asserting false statements, other forms of deception rely on the hearer reaching a conclusion from what the agent *implies.*

In more detail, intentional (non-lying) misleading involves asserting or stating something that is true but intending that the hearer infers from it something false, which the hearer thereby comes to believe. Importantly, there is no lie told – all statements are true, but the misleader intends that the listener will go on to infer something that is false.

There are many ways that this can be achieved within medical practice. One common method is by *non-disclosure* – the doctor might simply not answer the patient's question or not provide information that is clinically indicated. Alternatively, the doctor might answer a different question which the patient did not ask. Or the doctor may *deliberately misunderstand* the question, and hence confuse the patient.

Or the doctor may provide a *partial answer* to the patient's question – missing out the most unpalatable part of the answer.

Alternatively, a doctor may deceive by *euphemism* such as when a doctor says about a patient's heart failure that *'there is a bit of fluid on the lungs as the heart is not quite pumping strongly enough'* (Cox and Fritz (2015, p632). Again, this is actually a true statement, but use of euphemism downplays the seriousness of the problem and hence misleads the patient. Doctors often *blind the patient with science* – they provide a mass of true (but irrelevant) data which confuses the patient and leads to the patient reaching a false conclusion. Sometimes health care professionals *equivocate* and give a range of diagnostic possibilities, all of which are *possibly* true, but contain the most likely diagnosis. The patient may latch on to the most benign of the diagnoses offered, even if this particular diagnosis is extremely unlikely to be true.

Non-lying deception does not even need to be verbal. Some hospitals employ a 'slow code' where the patient is resuscitated but the resuscitation efforts are sub-optimal. The patient infers from the appearance of the resuscitation team that 'everything has been done'.

The key features of all these methods of deception is that a true statement is made each time, but it is made in such a way as to mislead the recipient of the (verbal or non-verbal) utterance.

Having set out the distinction between two methods of deception (lying v non-lying) I next turn to a consideration of whether these two methods differ *morally*. There is a large literature on this subject, and strong arguments have been put forward in both directions – claiming either that there is not, or that there is, a moral difference

between the two ways of deceiving. In the next two sections I summarise and critique the main arguments in favour of both views (Section 3 and 4). The aim of these two sections is to show that there are problems with arguments in both directions. This allows me (Section 5) to present my claim that there is a moral difference between lying and non-lying methods of deception, and that this moral difference is derived from the moral difference between doing and allowing harm.

6.4.3 Arguments that there is no moral difference between lying and non-lying forms of deception

(i) Roger Higgs's 'intention-only' account.

Higgs offers an 'intention-only' account. Higgs argues that in determining the moral rightness or wrongness of an attempted deceit, the only thing that matters is the intentions with which the deception is attempted.

> '*Surely it is the intention that is all important. We may be silent, tactful or reserved, but if we intend to deceive, what we are doing is tantamount to lying. It is hard but vital to see one's own evasion, duplicity or equivocation for what it is – a lie'. (Higgs, 1985, p189).*

Furthermore, says Higgs, there is no important moral difference between lying and other forms of intentional deception - that it depends not on how the deception is achieved, but only on the intention to deceive.

> *'it does not matter morally whether deception is achieved with a lie or with equivocation, evasion, by being 'economical with the truth', or merely from refraining from correcting a misunderstanding. It is of course, quite possible, and very common in clinical practice, for doctors or nurses to engage in deliberate deceit without actually saying anything they believe to be false....*

> *But....one could hardly argue that this was morally any different from telling outright lies'. (Higgs, 1985, p189).*

Higgs's account correctly captures the importance of intention in making moral assessments of deception. If the intention is the same, (and the outcome is the same) then it easy to see that the mechanism of deception is not important. It seems plausible, however that other moral factors matter too. In my account of the moral difference between doing and allowing harm, I argue that *causation* is also important in the allocation of blameworthiness for harmful outcomes. Whether the harm was caused or not, (I argue) makes a difference even if the intention is the same. Although, as I discuss in my analysis of the morality of the 'Bathtub' case, the causation component can be dwarfed (and perhaps not noticed) if the intention is egregious enough.

(ii) Jennifer Saul's 'same motivation/same result' account.

Saul, like Higgs, thinks it strange that we should prefer one method of deception to another, when the outcome and the motivation is the same. Her main argument relies on appeals to intuition, based on comparison of cases of lying and mere misleading. Consider Frieda:

> *Frieda suffers from a peanut allergy so dire that even the tiniest amount of peanut oil could be deadly. George knows that, and invites her to dinner, murderously preparing a stir-fry with peanut oil. Frieda, appropriately cautious, (though not cautious enough) asks, 'Are there any peanuts in the meal?', to which George replies, 'No there are no peanuts'. Frieda eats the stir-fry and dies. (Saul, 2012, p3).*

What George does is not lying (there are no *actual* peanuts in the meal) and the anaphylactic reaction results from the presence of peanut *oil*, not peanuts. Saul plausibly asserts that this is no better that telling the outright lie: '*The meal is perfectly safe for you*' or '*There is nothing in this meal that could possibly cause you harm*'.

Saul does not just rely on intuitive claims; after all the story of Frieda and George might just be an exceptional case which cannot be generalised to all cases of lying and misleading. To bolster her argument, Saul takes aim at attempts to generate an explanation for the moral difference between lying and misleading. Her main target is the (frequently cited) claim that non-lying deception shifts the responsibility for deception away from the perpetrator and towards the victim[89]. Because the audience has to make inferences from a true statement to a false conclusion, some of the responsibility for non-lying deception belongs to the victim; for a lie, the entire responsibility lies with the liar.

Saul denies that sharing moral responsibility for a deception alters the badness of the deception. She asks us to consider two mugging victims; one is mugged in a well-lit safe part of town, with his money concealed. The other is mugged in a dangerous area of town, with money hanging out of his pocket. Clearly the second muggee is partly responsible for being mugged – but (says Saul) this does not alter the wrongness of the act of mugging, nor alter the blameworthiness of the mugger. Hence, Saul claims, the shared responsibility model is not correct.

I agree with Saul that the 'shared responsibility' account does not capture the difference between lying and non-lying forms of deception. But this does not rule out

[89] Such an argument was first advanced by MacIntyre (1994) but also discussed by Adler (1997) and Green (2001).

alternative accounts and other explanations for the moral difference between different forms of deception.

(iii) Bernard Williams 'same effect on trust' account

Williams (2002) denies that there is any moral difference between lying and non-lying means of deception, because he thinks that they both have the same impact on trust, which, for Williams, is the reason that deception is morally problematic. Williams thinks that we ''fetishise lying - by '*gratuitously treating the act of making a false assertion as morally different than more circuitous deceptive acts'* (1994, p337). For Williams, both lying and non-lying deception are wrong because they undermine trust.

I would agree that both methods have a negative impact on trust. I would argue, however, that (at least sometimes) the impact on trust of a lie is greater than a non-lying deception (perhaps that is a reason why we try our best to avoid lying and settle for a non-lying deception). Later, in Section 5, I provide a plausible explanation for this possible difference.

6.4.4 Arguments in favour of there being some moral difference between deceiving and non-disclosure (lying is worse than non-disclosure).

I now turn to arguments that (contrary to the claims in the previous section) suggest that there is a moral difference between lying and other methods of deception. This is actually my position, but I provide different explanations for this moral difference.

(i) Jennifer Jackson – a duty not to lie but no duty to not mislead

Jennifer Jackson (1991) argues that doctors (and members of the public) have a duty to avoid lying, but have no duty to avoid misleading. Jackson thinks avoiding

lying is important for the preservation of trust, particularly in the context of a doctor-patient relationship. But she thinks that this is less important for other forms of misleading, and gives example from day-to-day life to show that a) misleading is common and b) reasonably well accepted. She gives the example of the widespread acceptance of cosmetic use (and presumably cosmetic surgery) in our society - where the wearer of make-up (presumably) aims to deceive friends and colleagues about his or her attractiveness. I think there are problems with this argument. Firstly, it not clear that wearers of make-up aim to deceive – it may be that they merely prefer their appearance when wearing make-up and it is this (rather than an intention to deceive) that motivates the use of cosmetics. But perhaps more importantly, I am not convinced that this example translates to the medical situation, where there is a moral imperative for the patient to fully understand their condition (and it would be a failure of the doctor to not ensure that this has been achieved). Both lying and non-lying deception seem to equally prevent this understanding.

(ii) **Chisholm and Feehan – asymmetrical violation of trust.**

Chisholm and Feehan (1977), contra Williams, argue that lying damages trust more than misleading. To support this argument, they argue that the listener has a right to expect that the speaker believes what has been said is true but there is (conversely) no a right to expect this when it is implied or omitted. So (p153) they say:

> '*why is lying thought to be worse, other things being equal, than other kinds of intended deception....it is assumed that if a person L asserts a proposition p to another person D, then D has the right to expect that L himself believes p. And... if [L] asserts p to D, while believing himself that p is not true, then he violates this right of D*.

This seems plausible but why should this violation of trust not occur if L were to only imply something. Here Chisholm and Feehan give an example taken from Kant – if a person is packing a bag, then this does not give the friends of this person the right to assume he is about to make a journey. Maybe this is true in the packing example. But it does not seem to be the case in other implicatures. Consider this conversation between doctor and patient, where the patient has just had some blood tests which indicate a diagnosis of cancer.

Patient: Is there any abnormality in my blood tests?

Doctor: You will be OK to go on holiday next week.

There is no lie here (the patient is indeed OK to go on holiday next week) but the patient may assume that there is no serious problem. The patient has asked about blood tests, and the doctor has replied with a comment about holidays. Patients (indeed all persons involved in conversation) assume that comments made to them are relevant, and will try to preserve the assumption that the doctor is being co-operative. So the patient will take the doctor as saying there is nothing wrong with the blood tests because it is OK to go on holiday. So Chisholm and Feehan's claim does not go through – expectations apply to implicature as well as assertions. The patient has the right to expect the speaker to believe what they imply; not just what they say.

(iii) Strudler – lying causes a greater collapse of trust then simple misleading

Alan Strudler (2010) argued that lying damages trust ***more*** than simply misleading because we are able to carry on a productive conversation even after the misleading

has been discovered and this would not be the case with a lying. Consider the example I quoted earlier:

Patient: Is there any abnormality in my blood tests?

Doctor: You will be OK to go on holiday next week.

If the doctor had said – 'No, your tests are fine' (a lie) and this was discovered, then the trust between the patient and the doctor would collapse completely. However, if the doctor has used his '*OK to go on holiday'* ploy, and the patient has questioned this ('*What do you mean by talking about a holiday...I asked you about the tests*') then the conversation could still go ahead in a meaningful way without a complete loss of trust. Alternatively, the misleading might succeed – the patient might simply say '*That's great news'*.

This 'collapse of trust' argument may sometimes work, but it does not seem to me that this will always be decisive – the patient might feel that trust has also been lost if (for example) the doctor does not seem interested in the patient's problem or seems to be equivocating about a very serious matter.

Jonathan Webber (2013 p 651) offers a similar view to Strudler. He claims that lying damages both credibility in assertion *and* credibility in implicature[90], whereas misleading damages credibility in implicature. One can preserve credibility in assertion through speech that damages one's credibility in implication, but the converse is not true. If the patient questions the doctor's '*You will be OK to go on holiday next week'* statement the patient may feel that the doctor is attempting some form of non-lying deception – but still believe that the doctor would not lie. Whereas if the patient detects that the doctor is lying, then they will believe that the doctor is

[90] Webber uses technical language to differentiate the two types of deception I have been discussing above. Webber calls lying, deception by assertion, and he calls non-lying deception, deception by implication.

lying to them and might also attempt to mislead by non-disclosure or some other non-lying method of deception.

Strudler and Webber's claims are plausible, but it is not clear to me that these differences translate into *moral* differences.

6.4.5 My account of the moral difference between lying and non-lying deception

We seem, therefore, to have reached an impasse. On the one hand, there is some intuitive attractiveness in the idea that lying is morally worse than mere misleading; many of us would go out of our way to merely mislead rather than lie. This strategy has a long history; there is a story about St. Athanasius (296-373) who escaped pursuers by answering the question "*Where is the traitor Athanasius?*" with the true (but misleading) answer – "*Not far away*". On the other hand, Saul's story (2012, p3) of Frieda and the peanut oil is also compelling; how could there be a moral difference when the outcome (death) and the intention (causing death) are identical?

I now want to clarify these questions by applying my account of doing and allowing harm to the question of lying versus mere misleading.

Recall that my account of intentionally doing harm involves both intending harm and genuinely causing harm. By contrast, intentionally allowing harm involves intending harm, together with a causal explanation of that harm – but with no element of genuine causation. My suggestion is that there is an analogy both between lying and doing harm, and between non-lying deception and allowing harm. This analogy is not exact, and becomes more apparent when considering non-disclosure as a method of non-lying deception. Here is how the analogy might work:

When an agent lies, the agent intends that the hearer should develop false beliefs. But the agent also *causes* these false beliefs. A false idea is introduced and causes the hearer to develop a particular belief. Counterfactually, without the lie, the false belief does not develop. If a doctor tells a patient (falsely) "*You do not have cancer*", then the patient's belief ("*I do not have cancer*") is both intended and caused by what the doctor says.

By contrast, when an agent does not lie, but simply infers a falsehood, the agent intends the hearer to reach a false belief, but does not genuinely cause it. If a doctor does not mention the diagnosis of cancer to a patient, or simply states "*You will be OK to go on holiday next week*", then the doctor has not genuinely caused the false belief – this process has occurred by the efforts of the patient, or perhaps the friends and family of the doctor ("*He said I was OK to go on holiday; I must be alright*"). There is a causal explanation here ('had the doctor not talked about the holiday, then the false belief would not have occurred') but no genuine causation. There is an analogy here with my previous claim about euthanasia and withholding/withdrawing life supporting treatment (LST). In euthanasia, the lethal injection causes the patient's death, whereas in withholding/withdrawing LST the cause of death is the underlying illness.

I thus claim that the moral distinction between lying, and non-lying form of deception *are of the same kind* as the moral distinction between doing and allowing harm discussed earlier in this chapter.

6.4.6 Objections to this account

I now want to consider three objections to my account.

(i) Is this account not the same as the shared responsibility account?

There are obvious analogies between my account (based on intention/causation) and the shared responsibility account of MacIntyre (1994) and Green (2001) discussed earlier. Recall Saul's powerful argument, based on the story of the careful and reckless victim, that sharing the responsibility for a wrongful act does not dilute the wrongness of that act – both muggings are equally wrong, and we do not exonerate the mugger who claims, "*I only mug reckless tourists*". It might therefore be claimed that my intention/causation account is analogous to the shared responsibility account, and therefore vulnerable to the same sort of criticism. What does it matter if the victim of the deception partly (partly) causes the deception – surely it is the wrongness of the deception that is crucial?

I do not think that this criticism hits home. As I hope I have shown earlier in this thesis, the difference between genuine causation and mere causal explanation does matter for the ascription of wrongness and moral responsibility. The causal process for a lie and a mere deception are different, even if the intention is the same. So Saul's criticism of the shared responsibility account (that sharing responsibility does not dilute the wrongness of the act) does not gain traction against my intention/causation account.

(ii) Gradations of causation

It might also be argued that whereas my account works well for cases of non-disclosure, it is less plausible for other methods of non-lying deception, such as the use of euphemism. Here it seems reasonable to claim that the doctor was involved in some way, in causing the false belief (albeit, to a lesser degree than if the doctor had lied). I think that there are several responses to this criticism. One approach would be to argue that my account only applies to the distinction between lying and non-disclosure, and therefore my account does not provide an explanation for other

types of non-lying deception than non-disclosure. A second response would be to accept the criticism, and say that there is indeed a gradation in the wrongness of different types of non-lying deception; that non-disclosure is more wrong than euphemism, say.

(iii) Back to Frieda and the peanut oil

Intuitively, it does seem that the non-lying deception of Frieda about the peanut oil is just as bad as telling an outright lie. This seems to undermine my account which suggests that there would be *some* difference (however small) in the wrongness of the two sorts of deception. One way of dealing with this is to argue that for some cases (and the Frieda case would be a good example) the misleading is maximally bad – so wicked in fact, that lying could be morally no worse. For these sorts of cases, intention is everything, and adding a causation element would not make any moral difference.

A similar argument relies on the argument that health care professionals have a duty that patients fully understand their condition (and hence can make autonomous decisions about management). Here again, intention is everything, and 'trumps' any small moral difference between genuine causation and causal explanation.

6.4.7 Conclusion

The aim of this discussion has been to provide a third example of the practical relevance of my analysis of doing and allowing in cases in medical ethics. Thus I have argued that in issues related to lying/misleading, there is a moral difference between lying and non-lying deception – and this difference is explained by my account of causal and teleological explanation.

6.5 Chapter Conclusion

I have chosen these three topics (end of life care, termination of pregnancy, and lying/misleading) because I wanted to test whether my theoretical account of causal/teleological explanation has the resources to deal with practical dilemmas that are topic and controversial. I wanted to check if my preferred account could be applied to these tricky practical situations, and whether the conclusions were plausible and 'fell out of' the theoretical account without much ad hoc additions to the theory. I feel that both these desiderata are satisfied in these three areas – they produce conclusions that are realistic and are able to guide decision making.

CHAPTER SEVEN – CONCLUSION

In this thesis, I have attempted to provide answers to a number of questions related to the moral implications of doing and allowing harm.

In this conclusion I want to first of all identify some general principles that have emerged from my work, and then summarise each chapter in turn.

7.1 Overall conclusions

I have argued that there is a moral difference between intentionally doing and intentionally allowing harm. Thus my thesis gives some support for the Doctrine of Doing and Allowing (DDA). It does not provide unreserved support however - because I argued that in some circumstance there was *no* moral difference between doing and allowing harm. For example, as discussed in Chapter 5, some allowings of harm are so egregious that agents who allow harm or harms are just as blameworthy as agents who actually do the same harm. It thus appears that sometimes (but not always) there is a moral difference between intentional doing and allowing of harm.

Attribution of moral responsibility for doing and allowing harm is not a simple matter. At least two inter-related moral principles are in play. Part of the blameworthiness for an intentional doing or allowing of harm is related to the way that agents *cause* that harm. This is different for doings and allowings. Doings involve 'genuine' causation – the agent who does harm actually causes that harm. By contrast, agents who

allow harm do not cause harm (they do not cause anything). However, agents who allow harm are still related to the causal process; they causally explain the harm by their allowing.

A second factor that contributes to moral responsibility for harm is intention. As the name implies, intention is involved in both intentionally doing and allowing harm. On way of understanding the blameworthiness of an agent's intention is to consider the teleological sequence that the intention is part of – in particular the agent's overall *telos* or aim.

I argue that the overall moral responsibility for both doing and allowing of harm is derived from both a causal explanation and a teleological explanation.

I also show that the same principles apply to doing and allowing good (benefit) – this too can only be understood using a causal and a teleological explanation.

I argue that these general principles have practical application in day-to-day medical ethics decision making. In particular, they apply to several life and death dilemmas, including euthanasia, withholding and withdrawing life sustaining treatment, and late termination of pregnancy. They also help in the understanding of the moral difference between lying and allowing the patient to reach a false conclusion. It is encouraging that the theoretical principles derived in Chapters 1-6, are able to be applied to practical (medical ethics) scenarios and produce plausible outcomes and decisions.

7.2 Chapter by Chapter review

I will now review the main components of my argument in support of the DDA by summarising each chapter in turn.

In Chapter 1, as well as providing an introduction for the whole thesis, I emphasised the importance of considering the difference between doing and allowing, rather than the difference between acts and omission (which feature in some accounts of the DDA). I also stressed that allowing harm involves a failure to prevent harm - what seems to be crucial about allowing harm is that the agent could have prevented something bad from happening – but didn't.

In Chapter 2, I defended the DDA against one of its strongest source of criticism – the so-called Contrast Strategy, and in particular the 'Bathtub' case of James Rachels. I analysed several objections to the Contrast Strategy, and concluded that only much of the contrast strategy is sound. However, Rachels' claim that the results of a single contrast case can be generalised to all cases of doing and allowing harm is problematic, and I argue that this undermines his overall position. So it seems that *sometimes* there is no moral difference between doing and allowing harm (in that sense, Rachels is correct), but in other situations there is a moral difference between the two behaviours (in this Rachels was incorrect).

In Chapter 3 I argued that there is a difference between doing and allowing harm which can be understood in terms of a causal explanation. Doings, I suggest, genuinely cause harm and because of this causal role, agents are morally responsible for doings. Allowings, I argue at length, do not and cannot cause harm. However, allowings are part of the causal explanations for harm (even though they do not actually genuinely cause) and this can also explain some of the moral responsibility for the harm.

In Chapter 4 I move from cause to explanation, and in particular the 'Four Explanations' of Aristotle. This is a bridging chapter between Chapter Three and Chapter Five. I move from causal explanation to teleological explanation and suggest that teleological explanation may also have a role in explaining the moral responsibility for doing and allowing harm.

In Chapter 5 I argue for the role of intention in assessing the wrongness of doing and allowing harm. I argue against one philosophical trend to attempt to reduce (and maybe eliminate) the role of intention in the ascription of moral responsibility for actions. I respond to arguments from James Rachels and Gerhard Overland by developing a richer account of actions, more akin to Audi's account of conduct, to show that intention can alter the moral valence of an action. I also rely on the teleological sequence account of Elizabeth Anscombe to show that we cannot fully understand allowings or doing until we have a better appreciation of the telos that an agent aims at when they do or allow harm. In Anscombe's 'pumping man' story, we can only understand the moving of the man's arm up and down when we discover the intention of the man. Why is he moving his arm up and down? – in order to poison the water.

The intentional order is also the order of the agent's practical deliberation. It is then easy to see that the rightness or wrongness of the final aim determines the rightness or wrongness of what the agent is doing.

In Chapter 6, I apply the theoretical conclusions reached in the previous chapters of this thesis to practical issues within medical ethics. I demonstrate that the theoretical principles apply nicely to a range of practical dilemmas, such as the issue of

euthanasia v withholding life sustaining treatment, the ethics of late termination when the mother's life is at risk, and the moral difference between lying and withholding relevant information (that then allows the patient to reach the wrong conclusion).

7.3 Final comment

Finally, I hope to have provided an answer to Norcross's task, which I introduced in Chapter One of the thesis.

Norcross asked the ethical theorist to respond to the following challenge (Norcross (2003) p457):

> *It is always possible to object to an argument for a pair of examples that a particular distinction is sometimes morally relevant and sometimes not. The objection by itself, though, has no force. It must be accompanied by an explanation of why the distinction is morally relevant in some cases and not in others. This is much harder to provide. At the very least, then, Rachels' argument puts the burden of proof squarely on those who would invest the distinction between killing and letting die with moral significance.*

It is my hope that my account of causal and teleological explanation can go some way to meeting Norcross's challenge.

CHAPTER EIGHT - BIBLIOGRAPHY

Adams, F. and Mele, A. 1989. The Role of Intention in Intentional Action. *Canadian Journal of Philosophy* 19(4) 511-531

Adler, JE. 1997. Lying, deceiving, or falsely implicating. *The Journal of Philosophy* 94: 435-452

Adler, JE. 2018. Lying and Misleading: A Moral *Difference*. In: *Lying: Language, Knowledge, Ethics and Politics* ed Michaelson, E. and Stokke, A. pp301-318

Airedale NHS Trust v Bland 1993. AC:789

Annas, J. 1982. Aristotle on Inefficient Causes. *The Philosophical Quarterly*. 32(129): 311-326

Anscombe, GEM. 1963. *Intention*. Second Edition. Harvard University Press. Cambridge, Mass.

Anscombe, GEM. 1957. Intention. *Proceedings of the Aristotelian Society* 12(1), pp.101-106.

Anscombe, GEM. 1961. "War and Murder," in Stein, W. (ed). *Nuclear Weapons: A Catholic Response*. London. Merlin: pp.43–62.

Anscombe, GEM. 1979. Under a Description. *Noûs*, 13 (2) pp. 219-233

Aquinas, T. 1947 *Summa Theologica* trans Fathers of the English Dominican Province 1947.

Aristotle. 2008. *Physics*. Translated: Waterfield, R. Oxford World Classics. Oxford: Oxford University Press

Aristotle. 1953. *The Nicomachean Ethics*. Harmondsworth. Penguin

Armstrong, DM. 2014. The Open Door: Counterfactual v Singularist Theories of Causation. In *Causation and the Laws of Nature*. Ed. Sankey H. Dordrecht. Kluvier.

Asscher, J. 2007. Killing and Letting Die: The Similarity Criterion. *Journal of Applied Philosophy,* 24(3): 271-282

Asscher, J. 2008. The Moral Distinction Between Killing and Letting Die in Medical Cases. *Bioethics* 22:5 278-285

Audi, R. 2016. *Reasons, Rights and Values*. Oxford: Oxford University Press

Bach, K. 2010. Refraining, omitting and negative acts. In: O'Connor T, Sandis C. eds. *A Companion to the Philosophy of Action*. Chichester. Wiley-Blackwell, pp.50-58

Barnett, AG, Van de Pols, JC, Dobson, AJ. 2005. Regression to the Mean: What is it, and how to deal with it. *International Journal of Epidemiology*, 34: 215-220

Barry C, Lindauer M, Overland G. 2014. Doing, allowing and enabling harm, an empirical investigation. In Knobe J, Lombrozo T, Nichols S, (eds) *Oxford Studies in Experimental Philosophy,* Oxford: Oxford University Press.

Barry, C and Overland, G. 2016. *Responding to Global Poverty: Harm Responsibility, and Agency*. Cambridge. Cambridge University Press.

Bogen, J, 1974, Moravcsik on Explanation'. *Synthese*, 28: 19–25.

Beebee, H. 2004. Causing and Nothingness' in L. A. Paul, E. J. Hall and J. Collins, eds., *Causation and Counterfactuals*, MIT Press, Cambridge.

Benn, P. 2001. Medicine, Lies and Deception. *Journal of Medical Ethics* 27: 130-134

Bennett J. 1981. *The Act Itself.* Oxford: Clarendon Press.

Bennett, J. 1966. Whatever the Consequences. *Analysis* 26(3): 83-102

Bernstein, S. 2014. Omissions as Possibilities. *Philosophical Studies.* 167:1-23

Bostyn, DH, Roets, A. 2016. The Morality of Actions: The Asymmetry Between Judgement of Praise and Blame in the Action-omission Effect. *Journal of Experimental Social Psychology* 63: 19-25

Bratman, M. 1987. *Intention, Plans and Practical Reason*. Cambridge, MA: Harvard University Press.

Broadbent, A. 2013. Explanation and Responsibility. In: *Critical essays on Causation and Responsibility* ed Kahmen, B and Stephians M. Berlin: De Gruyter

Brock, D 1985 Taking human life. *Ethics* 95: 851-865

Brock, D. 1999. Voluntary active euthanasia in Beauchamp and Walters (ed) *Contemporary issues in bioethics*. Wadsworth. p296-305

Callahan, D. 2012. Killing and Allowing to Die: Why it is a Mistake to Derive an 'Is' from an 'Ought. In *The Roots of Bioethics: Health, Progress, Technology, Death*. Ed: Callahan D. pp 116-123

Cameron, R. 2010 Aristotle's Teleology. *Philosophy Compass* 5(12): 1096-1106

Carson, TL. 2006. The Definition of Lying. *Nous* 40:2 284-306

Charlton, W. 1992. *Aristotle's Physics Book I and II.* Oxford: Clarendon.

Coope, Ursula 2007. Aristotle on Action. *Proceedings of the Aristotelian Society Supp.* Vol. 81/1: 109– 38.

Chisholm, RM and Feehan, TD. 1977. The Intent to Deceive. *The Journal of Philosophy* 74(3): 143-159

Chisholm, R (1966) Freedom and Action. In Lehrer, K (ed) *Freedom and Determinism*. New York. Random House.

Clarke, R. 2014. *Omissions: Agency, Metaphysics, and Responsibility*. Oxford. Oxford University Press.

Colombo, C.F. 2018. *The doing/allowing distinction: causal relevance and moral significance*. PhD thesis, The London School of Economics and Political Science (LSE).

Cooper, John. 2013. Aristotelian Responsibility, *Oxford Studies in Ancient Philosophy Vol. XLV*, ed. Brad Inwood, Oxford: Oxford University Press: 265–312

Cox, CL and Frit, Z. 2016. Should Non-Disclosures be considered as morally equivalent to lies within the doctor-patient relationship? *Journal of Medical Ethics* 42: 632-635

Dancy, J. 2004. *Ethics without Principles*. Oxford: Oxford University Press.

Davidson, D 1967. Causal Relations. *The Journal of Philosophy*, 64:691-703

Davidson, D. 1963. Actions, Reasons and Causes, in: *Essays on Actions and Events, 3-20*. Oxford: Clarendon Press.

Davis, WA. 2010. Casual theory of action. In: O'Connor T, Sandis C. eds. *A Companion to the Philosophy of Action*. Chichester. Wiley-Blackwell, pp.32-40

Dinello, D. 1971. On Killing and Letting Die. *Analysis* 31(3): 83-86

Donagan, A. 1977. *The Theory of Morality.* Chicago: The University of Chicago Press.

Dowe, P (2000. *Physical causation.* Cambridge. Cambridge University Press

Dowe, P. 2001. A counterfactual theory of prevention and 'causation' by omission. *Australasian Journal of Philosophy.* 79(2), pp 216-226.

Draper, K. 2005. Rights and the Doctrine of Doing and Allowing. *Philosophy and Public affairs,* 33 (3): 253-280

Driver, J. 2008. Attributions of Causation and Moral Responsibility. In: Sinnott-Armstrong W. (ed) *Moral Psychology Volume 2 The Cognitive Science of Morality: Intuition and Diversity,* pp 423-441

Driver, J. 2018. *Gertrude Elizabeth Margaret Anscombe* [Online] [Accessed 1st December 2021] in Zalta, E.N. (ed.) 2011. The Stanford Encyclopedia of Philosophy. Available from: http://plato.stanford.edu/archives/win2011/entries/anscombe/

Dunaway, D, Edmonds A, Manley ,D. 2013. The Folk probably do think what you think they think. *Australasian Journal of Philosophy*. 91(3): 421-44

Ehring, D. 2009. Casual relata. In: *The Oxford Handbook of Causation*. Eds Beebee H, Hitchcock C, Menzies P. Oxford. Oxford University Press.

Elster, J. 2015. Doing and Allowing Harm: Review. *Notre Dame Philosophical Review.* https://ndpr.nd.edu/reviews/doibg-and-allowing-harm/

Falcon, A. 2023. Aristotle on Causality, *The Stanford Encyclopedia of Philosophy* (Spring 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = https://plato.stanford.edu/archives/spr2023/entries/aristotle-causality/

Foot, P. 1967. The Problem of Abortion and the Doctrine of Double Effect. *Oxford Review* 5, pp 1-6

Foot, P. 1985. Morality, Action and Outcome: in T Hondereich (ed) *Morality and Objectivity*. London: Routledge.

Foot, P. 1994. Killing and Letting Die. In Steinbock, B and Norcross, A (eds) *Killing and Letting Die*, 2nd Edition, New York: Fordham University Press. pp 280-289

Francis, R. 2013. *Report of the mid-Staffordshire NHS Foundation Trust Public Enquiry.*

Frey, J. 2022. Anscombe on the voluntary character of negligence. In *Teleological structures in human life: Essays in honour of Anselm J Muller*. pp26-42 ed Christian Kietzmann.

Garrard, E and Wilkinson S. 2005. Passive euthanasia. *Journal of Medical ethics.* 31: 64-68

Gettier, EL. 1963. Is Justified True Belief Knowledge? *Analysis*, 23(6): 121–123

Green, SP. 2001. Lying, misleading, and falsely denying: how moral concepts inform the law or perjury, fraud and false statements. *Hastings Law Journal* 53: 157-212.

Haas, JM. 2017. Moral theological analysis of direct versus indirect abortion. *The Linacre Quarterly* 84(3): 248-260

Hall, N. 2004. Two Concepts of Causation in Collins J, Hall N, and Paul LA (eds) *Causing and Counterfactuals*. Cambridge, Mass. MIT Press (pp 225-277)

Hennig, B., 2009, “The Four Causes” *Journal of Philosophy*, 106: 137–160.

Herring, J. *Great Debates in Criminal Law*. Fourth Edition. Oxford: Bloomsbury Publishing.

Higgs, R. 1985. On Telling Patients the Truth. In: *Moral Dilemmas in Modern Medicine*, ed Michael Lockwood pp187-203. Oxford: Oxford University Press.

Hill, S. 2018. Murdering an Accident Victim: A New Objection to Rachels’ Bare-Difference Argument. *Australasian Journal of Philosophy* 96(4): 767-78

Hitchcock, C. 2021 "Probabilistic Causation", *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition), Edward N. Zalta (ed.), URL <https://plato.stanford.edu/archives/spr2021/entries/causation-probabilistic/>. Accessed 1st August 2022.

Hocutt, M. 1974. Aristotle’s Four Becauses. *Philosophy* 49: 385-399

Hume, D. 1888. *A Treatise of Human Nature,* ed Selby-Bigge, LA. Oxford: Clarendon Press

Illari, P. 2010. Why theories of causality need production: an information-transmission account. Philosophy and Technology 24(2): 95-114

Illari, P and Russo, F. 2014. *Causality: Philosophical Theory meets Scientific Practice*. Oxford: Oxford University Press.

Jackson, F. 1982. Epiphenomenal Qualia. *Philosophical Quarterly* 32: 127-136

Jackson, J. 1991. Telling the Truth. *Journal of Medical Ethics*. 17: 5-9.

Johannson ,J and Risberg, OP. 2022. A Simple Analysis of Harm. *Ergo: An Open Access Journal of Philosophy*. 9:509-536

Kagan S, 1989 *The Limits of Morality.* Oxford: Clarendon Press

Kagan, S, 1988. The Additive Fallacy. *Ethics*, 99 (1): 5-31

Kahneman, D. 2011. *Thinking Fast and Slow*. London: Penguin

Kamm, FM. 1983. Killing and letting die – Methodological and substantive issues. *Pacific Philosophy Quarterly* 64(4): 297-312

Kamm, FM. 2007. *Intricate Ethics. Rights, Responsibilities and Permissible Harm*. Oxford: Oxford University Press.

Kopeikin, Z.A. 2019. Bare-Difference Methodology and a Problematic Separability Principle. *The Journal of Value Enquiry* 54*,* pp.553-570

Kopeikin, Z.A. 2020. A Separability Principle, Contrast Cases, and Contributory Dispositions. *Southwest Philosophy Review.* 36(1), pp.35-44.

Lange, M. 2016. *Because without Cause: Non-causal explanation in Science and Mathematics.* Oxford. Oxford University Press

Moravcsik, J. M., 1974, Aristotle on Adequate Explanations. *Synthese*, 28: 3–17.

Lepora, C, and Goodwin RE. 2013. *On Complicity and Compromise*. Oxford: Oxford University Press.

Lewis, D. 1973. Causation. *Journal of Philosophy* 70: 556-567

Lewis, D. 1986. Causal Explanation. In *Philosophical Papers* Volume 1. Oxford. Oxford University Press.

Lewis, D. 2000. Causation as influence, in Collins J, Hall N, and Paul LA (eds) *Causing and Counterfactuals*. Cambridge, Mass. MIT Press (pp 75-107)

Lichtenberg, J. 1982. The Moral Equivalence of Action and Omission. Canadian Journal of Philosophy Supplementary Volume 8:19 – 36.

Livengood, J and Machery, E (2007). The Folk Probably Don't Think What You Think They Think: Experiments on Causation by Absence. *Midwest Studies in Philosophy* XXXI: 107-127

MacIntyre, A. 1994. Truthfulness, lies and moral philosophers: what can we learn from Mill and Kant? In *Tanner Lectures on Human Values* 309-369. Princeton: Princeton University Press.

Mackie. JL. 1974. The Cement of the Universe. Oxford. Oxford University Press

Malm, H. 1992. Defence of the Contrast Strategy. In: Fischer, J.M and Ravizza M. (eds.) *Ethics Problems and Principles*, Fort Worth: Harcourt, Brace, Jovanovich Publishers.

McGrath, S. 2003. Causation and the Making/Allowing Distinction. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 114: 81-86

McGrath, S. 2005. Causation by Omission: A Dilemma. *Philosophical Studi*es 123: 125-148

McLachlan, HV. 2008. The Ethics of killing and letting die: active and passive euthanasia. *Journal of Medical Ethics.* 34: 636-38.

McMahan, J. 1993 Killing, letting die, and withdrawing aid, *Ethics* 103: 250-279

Mele, A. 2010. Intention. In: O'Connor T, Sandis C. eds. *A Companion to the Philosophy of Action*. Chichester. Wiley-Blackwell, pp.108-11

Mellor, DH 1995. *The Facts of Causation*. London. Routledge.

Menzies, P and Beebee, H. 2020 Counterfactual Theories of Causation. *The Stanford Encyclopedia of Philosophy.* Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2020/entries/causation-counterfactual/>. Accessed 16th May 2022

Menzies, P. 2011. The Role of Counterfactual Dependence in Causal Judgments.' *Understanding Counterfactuals, Understanding Causation: Issues in Philosophy and Psychology*. Eds. Hoerl, C McCormack, T, and Beck SR. Oxford: Oxford University Press.

Miller, FG, Truog, RD, Brock, DW. 2008. *Death dying and organ transplantation; Reconstructing medical ethics at the end of life.* Oxford: Oxford University Press.

Miller, FG, Truog, RD, Brock, DW. 2010. Moral fictions and medical ethics. *Bioethics* 24 (9) 453-460

Mogensen, A. and MacAskill, W. 2021. The Paralysis Argument. *Philosopher's Imprint.* 21 (15):1-17

Montgomery (Appellant) v Lanarkshire Health Board (Respondent). 2015. UKSC: 11

Moore, MS. 2009. *Causation and Responsibility. An Essay in Law, Morals and Metaphysics.* Oxford. Oxford University Press.

Moravcsik, JME. 1974. Aristotle on Adequate Explanation. *Synthese* 28(1): 3-17

Mumford, S. 2021. *Absence and Nothing. The Philosophy of What There is Not.* Oxford. Oxford University Press.

Nesbitt, W. 1995. Is Killing No Worse Than Letting Die? *Journal of Applied Philosophy.*

Norcross, A. 2003. Killing and letting die. pp.451-463. *A Companion to Applied Ethics.* Oxford: Blackwell.

Overland, G. 2014. Moral obstacles: An Alternative to the Principle of Double Effect. *Ethics* 124 (3): 481-506

Perrett, RW. 1996. Killing, Letting Die, and the Bare Difference Argument. Bioethics 10(2): 131-139

Persson, J. 2002. Cause, Effect, and Fake Causation. *Synthese*, 131;129-143

Purves, D. 2011. Still in Hot Water: Doing, Allowing, and Rachels' Bathtub Cases. *Southwest Philosophy Review.* 27(1), pp.129-137.

Pust, J. 2019. Intuition. [Online] [Accessed June 21st, 2021].

Quick, O. 2022. Duties of Candour in Healthcare: The Truth, The Whole Truth, and Nothing but the Truth. *Medical Law Review*, 30(2): 324-347

Quinn, WS, 1989. Actions, Intentions and Consequences: The Doctrine of Doing and Allowing. *Philosophical Review*, 98(3):287-312

R v Cox. 1992. 12: *BMLR* 38.

Rachels, J. 1975. Active and Passive euthanasia. *New England Journal of Medicine* 292:78-86

Rachels, J. 1979. Killing and Starving to Death. *Philosophy* 54: 159-171

Rachels, J. 1981. Reasoning about Killing and Letting Die. *Southern Journal of Philosophy* 19(4), pp.465-473

Rachels, J. 1986. The End of Life: Euthanasia and Morality. Oxford. Oxford University Press.

Rasanen, J and Hayry, M. 2024. Fifty years of killing and letting die: On the limits of philosophical bioethics. *Bioethics*:, 1-5.

Re: B. 2002. Adult: refusal of medical treatment. *All England Law Bulletin*. 449.

Reiss, J. 2019. Causal Explanation is all there is to Causation. *Teorema* XXXVIII(3): 25-52

Reutlinger, A. and Saatsi J. 2017. *Explanation Beyond Causation: Philosophical Perspectives on Non-Causal Explanations.* Oxford. Oxford University Press.

Sandberg, J and Juth, N. 2011. Ethics and Intuitions: A Reply to Singer. *The Journal of Ethics 15 (3): 209-226*

Sartorio, C. 2007. Causation and Responsibility. *Philosophy Compass* 2: 749-765

Sartorio, C. 2009. Causation and Ethics. In Beebee, H Hitchcock, C Menzies, P. *The Oxford Handbook of Causation.* Oxford. Oxford University Press.

Sartorio, C. 2022a. Responsibility and Causation. In Nelkin DK Pereboom, D. *The Oxford Handbook of Moral Responsibility.* Oxford. Oxford University Press. (pp 548-365)

Sartorio, C. 2022b. Juggling Intuitions about Causation and Omissions. *Advances in Experimental Philosophy of Causation* eds Willemsen P and Wiegmann A. London: Bloomsbury

Saul, J. 2012. Just Go Ahead and Lie. *Analysis* 72 (1): 3-9

Savulescu J. 2013. Abortion, infanticide, and allowing babies to die – 40 years on. *Journal of Medical Ethics* 39(5): 257-259

Sehon, S. 2013. Teleological Explanation. In: O'Connor T, Sandis C. eds. *A Companion to the Philosophy of Action*. Chichester. Wiley-Blackwell, pp.121-129

Sehon, S. 2016. *Free Will and Action Explanation: A Non-causal Compatibilist Account.* Oxford. Oxford University Press.

Singer, P. 1993. *Practical Ethics,* Second Edition. Cambridge, Cambridge University Press.

Singer, P. 2005. Ethics and Intuitions. *The Journal of Ethics, 9:3/4 331-352*

Sinnott-Armstrong, W, Young, L, Cushman, F. 2010. Moral Intuition. In 'The Moral Psychology Handbook, ed Doris JM. Oxford. Oxford University Press

Steward, H. 1997. *The Ontology of Mind.* Oxford. Oxford University Press

Stratton-Lake, P 2002. *Ethical Intuitionism Re-evaluated*. Oxford: Oxford University Press p1

Strawson, PF. 1963 (2003). Reprinted in Gary Watson ed *Free Will*, Second Edition, pp72-93. Oxford: Oxford University Press.

Strawson, PF. 1992. *Analysis and Explanation in Metaphysics: An Introduction to Metaphysics.* New York: Oxford University Press. pp109-131

Strudler, A. 2010. The Distinctive Wrong in Lying. *Ethical Theory and Moral Practice.* 13(2): 171-179

Stuchlik, J. 2022. *Intention and Wrongdoing*. Cambridge. Cambridge University Press.

Tannsjo, T. 2005. *Taking Life.* Oxford: Oxford University Press

Taurek, J. 1977. Should the Numbers Count? *Philosophy and Public Affairs* 4:293-316

Thomson, JJ. 1971. A Defense of Abortion. *Philosophy and Public Affairs* (1): 47-66

Thomson, Judith Jarvis. 2003. Causation: Omissions. *Philosophy and Phenomenological Research* LXVI (1) pp 81-103

Thomson, Judith Jarvis. 1976. Philosophical Problems of Death. Killing, Letting Die and the Trolley Problem. *The Monist*, 59 (2), pp. 204-217

Thomson, Judith Jarvis. 1991. Self defense. *Philosophy and Public Affairs*, 20(4): 283-310

Tooley, M. 1972. Abortion and Infanticide. *Philosophy and Public Affairs* 2(1): 37-65

Tooley, M. 1994. An Irrelevant Consideration: Killing versus Letting Die. In Steinbock, B and Norcross, A (eds) *Killing and Letting Die*, 2nd Edition, New York: Fordham University Press. pp 103-112

Tracey v Cambridge University Hospital NHS Foundation Trust and another. 2014. *All ER (D) 138*

Trammell, RL. 1975. Saving life and taking life. *The Journal of Philosophy*, 72:131-137

Trammell, RL. 1976. Tooley's Moral Symmetry Principle. *Philosophy & Public Affair*s 5(3): 305-313

United States Conference of Catholic Bishops (USCCB) 2009. *Ethical and Religious Directives for Catholic Health Care Services 5th Edition*. Washington, DC: USCCB

Unruh, CF. 2022. Doing and Allowing Good. *Analysis* 82(49: 630-637.

Varzi, AC. 2007. Omissions and Causal Explanations. In: Castellani, F and Quitterer J (eds) *Agency and Causation in the Human Sciences*. Paderborn: Mentis Verlag, pp155-167

Vlastos, G. 1969. Reasons and Cause in the Phaedo. *The Philosophical Review* LXXVIII (3) p294

Webber, J. 2013. Liar! *Analysis* 73 (4): 651-659

Wedgwood, R. 2011. Defending Double Effect. *Ratio*, XXIV (4) 384-401

Wilkinson, D, Herring, J, Savulescu, J. 2020. *Medical Ethics and Law: A Curriculum for the 21st Century.* Edinburgh: Elsevier.

Williams, B. 2002. *Truth and Truthfulness*. Princeton: Princeton University Press

Wilson, G. 1989. *The Intentionality of Human Action*. Stanford, CA. Stanford University Press

Woodward, J and Ross L. 2021. Scientific Explanation, *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2021/entries/scientific-explanation/>.

Woollard, F. 2011. Most Ways I Could Move: Bennett's Act/Omission Distinction and the Behaviour Space. *Mind* 120: 155-182

Woollard, F. 2012. The doctrine of doing and allowing I: Analysis of the doing/allowing distinction. *Philosophy Compass.* 7(7), pp.448-458.

Woollard, F. 2012. The doctrine of doing and allowing II: The Moral Relevance of the Doing/allowing distinction. *Philosophy Compass.* 7(7), pp.459-469.

Woollard, F. 2013. If this is my body...: A Defence of the Doctrine of Doing and Allowing. *Pacific Philosophical Quarter*ly 94: 15-341

Woollard, F. 2015. *Doing and Allowing Harm.* Oxford: Oxford University Press

Woollard, F. 2016. Doing vs Allowing Harm *Stanford Encyclopaedia of Philosophy* https://plato.stanford.edu/entries/doing-allowing/

Woollard, F and Howard-Snyder, F. 2024. "Doing vs. Allowing Harm", *The Stanford Encyclopedia of Philosophy* (Winter 2024 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/win2024/entries/doing-allowing/>. Accessed 3rd March 2025.

Xess, S and Shrivastava, S (2019). Role of fetal craniotomy in modern day obstetrics: case series. *International Journal of Reproduction, Contraception, Obstetrics and Gynaecology* 8(4) 1679-1682