

Contingency, Tragedy and Politics

Kenta Sekine
UCL

Thesis submitted for the degree of PhD in Philosophy

I, Kenta Sekine, confirm that the work presented in my thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

29th March 2025

おじいちゃま、おばあちゃま。
こちらこそ、いつまでも遠くから応援しています。

Abstract

The question that guides this thesis is: how should we respond to contingency in our lives? I address this question as it arises for what I argue is the most troubling face of contingency in our lives, the phenomenon I call ‘tragedy.’ Tragedy befalls us, in this sense, when we find ourselves apt to experience feelings of moral responsibility for something we did, even though we were not at fault in so doing.

We can hear the guiding question in two registers. How should we each respond to tragedy? And how should we collectively respond?

At the individual level, I argue that tragedy proves recalcitrant to philosophical understanding unless we can account for a feeling of responsibility that is like agent-regret in being apt in the absence of fault, but like blame in being of a ‘moral’ quality sufficient to do justice to the phenomenology of tragedy. I propose to account for such a feeling in terms of ordinary agent-regrets that are hard to let go of, or repress, because they respond to failures to satisfy important obligations of ours. It is with such feeling that we should, as individuals, respond to tragedy in our lives. To defend this view, I offer accounts of ordinary agent-regret and the distinctive normative force of obligation.

With this in hand, I go on to address the guiding question at the level of the collective. I do so by bringing out what I believe to be the political dimensions of tragedy. I consider how the societal distribution of tragedy is affected by struggles over the obligations that in fact constitute the value of our social relations, and over how we collectively understand the role of that value in human life. We should collectively respond to tragedy, I claim, by engaging in such struggles.

Impact Statement

As of 2018, UCL requires all PhD theses to include an impact statement, which should describe how the expertise, knowledge, analysis, discovery or insight presented in the thesis could be put to a beneficial use.

This thesis lies at the intersection of moral, legal and political philosophy, and aims to make a scholarly contribution to these subfields. In moral philosophy, the thesis offers a novel account of a feeling of responsibility with a distinctively ‘moral’ quality that we may be genuinely apt to feel in the absence of personal fault. This opens up new ground in the theory of responsibility, agency and emotion, and may be useful to applied philosophers and social scientists interested in understanding how structural factors impact the emotional lives of individuals. The possibility of faultless moral responsibility may also be useful to legal theorists interested in understanding norms of strict liability, especially in private law. In later chapters, the thesis offers an account of the value of our ability to make contractual arrangements, and the potential ills of over-contractualising our social relations, which may be useful to legal theorists and policymakers interested in the moral costs of markets and outsourcing. In addition, the thesis offers, in a limited domain, an account of what makes certain considerations distinctively ‘political’ in the sense that has interested political ‘realists’ in their ongoing debate with the opposed ‘moralist’ camp, which may prompt further scholarship in that area. I aim to publish versions of this material in academic journals, so that the above benefits are more likely to be realised.

There is also some potential to foster interdisciplinary dialogue, particularly around the law and practice of social care in the UK. The discussion of Chapter Six addresses the human costs of one guise of the contractualisation of citizen-state relations, crystallised in the Care Act 2014 and related legislation, which may be of use to socio-legal theorists and social care practitioners (and indeed recipients) in the UK context.

Acknowledgements

When I started doing philosophy, I was a relative outsider to its institutions. I thought that I could become an academic and use that as a part-time job to fund what I had been doing before, which was music. I can't remember when it dawned on me how naïve that was. It has been a long and sometimes confusing journey to being relatively inside. In that time, my world has grown, I have grown, and I have been allowed to develop as a philosopher. I count myself very lucky to have had the privilege. There are many who I would like to thank.

First of all, I am grateful for the financial support of LAHP and UCL, which enabled me, amongst other things, to complete this research project.

Ulrike Heuer has been there to guide my thesis through every stage of its development. I imagine that this would, at times, have been a frustrating task. I can admit that I tend to want my work to express something like a worldview, and I worry that this sometimes leads me to instrumentalise my thinking. Ulrike never does this; she lets the arguments lead where they lead, as a philosopher should. I have a long way to go before I have anything like her courage. The more of it you have, the more of it you need. I am often reminded of a conversation we had in Regent's Park about some terrain in value theory, in which she described her fright about the problems there, because she could see what I could not see. Ulrike is a powerful, exacting thinker, and I respect her very much. I am so grateful for the time and effort she put into helping me think more earnestly through the ideas in this thesis.

Mark Kalderon was my secondary supervisor in the department, and he gave a valuable perspective on the ideas contained in several chapters of the thesis. My whole approach to philosophical research has been influenced by his advice and example: I learned from him to write better, to read voraciously, to resist the pressures of over-specialisation, and to try to say something new. Prince Saprai also supervised me during my year in the Faculty of Laws, and I thank him for generously continuing to meet with me even after

I left. The thesis would have been quite different were it not for our conversations about contract and private law theory, Gardner, and Raz, and it might not have existed at all without his encouragement and insistence on clarity about its overall structure. Nandi Theunissen supervised me in Pittsburgh, and I thank her for welcoming me into the city and for involving me in her fantastic seminar on Murdoch, and the ABD workshop too. Our conversations about the profession, and about what it is, or could be, to do philosophy left a lasting impression on me.

At UCL, I have been blessed with all the benefits that come with a vibrant departmental environment and culture, not least our graduate community. For all sorts of things over the years, I would like to thank Euan Allison, George Britten-Neish, Christabel Cane, Luka Chilvers, Lucas Chebib, Diana Craciun, Alice Harberd, Karri Heikkinen, Lizzy Holt, Tasnim Ismail, Kristine la Cour, Kim Leontiev, Harry Long, Alex Murphy, Timo Schobinger, Michael Thorne, Nikhil Venkatesh, Fiona Whittingham, Victor Braga Weber, and especially Jack Hume and Filippa Ronquist. Thanks to Victor for organising the political realism reading group, which kept me sane during lockdown, and Amanda Greene for inviting some wonderful guests, including Steven Lukes, which was a real treat. Thanks also to our departmental administrators, without whom it would all come tumbling down, and to Richard Edwards, for a very timely piece of advice.

I am grateful to my colleagues Edward Adegbemi, Lauren Cottrell, Joseph Henry, Peter Holzmann and Selsabil Benbellil at Southwark Libraries for their insights, and Eugenia Atta, Meno Jacob and Amy Houtenbos for accommodating me in various ways.

To my friends in London, Copenhagen and now Barcelona, thank you for all the times (good and bad) and to Ben, Callum, Cam, Colin and Sammy, who are always there.

Thanks to Gavin and Nettie for always taking an interest, and from whom I have learned a lot, and to my parents, Alan and Asuko, who, in a complicated way, taught me to think philosophically. Yuta, brother, it just goes without saying.

Nina, without you, I could not have made it through. There is nothing we cannot discuss, and that is so much of our strength. Blimey, we are many chapters deep, aren't we? I look forward to what the next holds.

Contents

Part One. *Contingency*

Preamble to Part One.....	11
---------------------------	----

Chapter One. Faces of Contingency in Our Lives

1.1. Contingency and Moral Luck.....	13
1.2. Some Preliminaries.....	16
1.3. Senses of ‘Control’.....	19
1.3.1. ‘Control’ as Reasons-Responsiveness.....	20
1.3.2. Failures of Receptivity and Reactivity.....	23
1.3.3. Justifications and Excuses.....	27
1.3.4. ‘Control’ as Faulty Reasons-Responsiveness.....	33
1.4. Tragedy and the Fault Principle.....	36
1.5. Concluding Remarks.....	40

Chapter Two. Against Moralism about Obligation

2.1. Everyday Tragedy.....	41
2.2. The Debate over ‘Moralism’.....	43
2.3. Blameworthiness Entails Fault.....	48
2.4. Breach Does <i>Not</i> Entail Blameworthiness.....	54
2.5. Interpreting Feelings of Responsibility.....	61
2.6. Concluding Remarks.....	65

Part Two. *Tragedy*

Preamble to Part Two.....	66
---------------------------	----

Chapter Three. Why Feel Agent-Regret?

3.1. Agent-Regret Eliminativism.....	78
3.2. Explaining Reasons for Action.....	82
3.3. Value, Attachment and Emotional Vulnerability.....	89
3.4. The Primacy of Activity in Life.....	96
3.5. Concluding Remarks.....	99

Chapter Four. On the Reasons We Cannot Escape

4.1. Preliminary Remarks.....	102
4.2. Respect, Engagement and Attitude-Independence.....	103
4.3. Attachment-Dependent Obligations.....	107
4.4. The Force of Inescapability.....	111
4.5. Concluding Remarks.....	117

Chapter Five. In Defence of Second-Order Reasons

5.1. What are Second-Order Reasons?	118
5.2. Understanding the Conceptual Challenge.....	121
5.2.1. Principle (LR)	122
5.2.2. The CREDIT ARGUMENT.....	126
5.3. The Right Reasons are Second-Order.....	127
5.4. The Right Reasons are <i>not</i> First-Order.....	133
5.5. Do the Examples Generalise?	135
5.6. Concluding Remarks.....	139

Part Three. *Politics*

Preamble to Part Three.....	141
-----------------------------	-----

Chapter Six. Contractualisation and its Discontents

6.1. What is Contractualisation?.....	144
6.2. A Moral-Philosophical Analysis.....	146
6.2.1. Contract as an Instrumental Relationship.....	147
6.2.2. Disposable Social Relations.....	153

6.3. Against Inflationism in Contract Theory.....	159
6.3.1. The <i>Doux Commerce</i> Thesis.....	161
6.3.1.A. Macrolevel Interpretations.....	164
6.3.1.B. Social-Relational Microlevel Interpretations.....	171
6.3.1.C. Dispositional Microlevel Interpretations.....	184
6.3.2. Pessimistic Inflationism.....	187
6.3.2.A. Clear Nonstarters.....	190
6.3.2.B. The Politics of Personal Detachment.....	197
6.3.3.C. Contract as a Form of Technology.....	204
6.4. Concluding Remarks.....	209
 <i>Chapter Seven. Security, Social Relations, and the ‘Political’</i>	
7.1. Our Security Interest.....	211
7.2. Local Distributions of Security.....	214
7.3. The Distinctively ‘Political’.....	220
7.4. Politics and Forms of Liability.....	227
7.5. Concluding Remarks.....	232
 Bibliography.....	 234

Part One

Contingency

Our lives, for better or for worse, are not totally under our control. Try as we might to avoid it, things can just happen—sometimes unavoidably—that push our lives off course. That is because the world may well be, and certainly may as well be, an endless expanse, one whose powerful machinations are to a large extent indifferent to our fates. The pages that follow address the question: how should we respond to the influence of contingency in our lives?

I think the question can be heard in at least two registers. The first is: how should we each respond to contingency in life? The second is: how should we, together, respond to contingency in life? It is a methodological assumption of this dissertation that if there are facts of the matter about the first question, then our answer to the second question should not ride roughshod over them. In this sense, I agree with Ronald Dworkin (2002: 323-4) that answers to the question at the level of the collective must be sensitive to our practices at the level of personal ethics, that they be in that sense “continuous” with

“assignments of responsibility drawn from ethics.” To that extent, but probably only to that extent, what follows can be considered an exercise in what he would call ‘continuous theory.’

Of course, my guiding question is an enormous, sprawling question—too much so to reasonably expect to answer in a serious way in a single dissertation. The point of Part One is to substantially narrow it down, bringing out the main difficulty of answering it in even its narrower form, and to gesture towards how we might overcome the difficulty. Chapter One, ‘Faces of Contingency in Our Lives,’ first narrows down the question to the impact of those life events beyond our control that do not merely happen to us, but in which we play an active role: life events that result from what Bernard Williams (1976) and Thomas Nagel (1976) famously called ‘moral luck.’ I then narrow the question down further to the form of moral luck often called ‘resultant luck,’ and yet further to the form of resultant luck I call *tragedy*. I suggest that responding to the impact of this particular face of contingency poses a unique problem, which orients the discussion to follow in Parts One and Two.

Chapter Two, ‘Against Moralism about Obligation,’ takes up the unique problem as it arises in a debate over the very coherence of tragedy in the sense explained in Chapter One. The debate, as I present it, is between those who deny the coherence of tragedy—a view I call ‘moralism’ about obligation—and two forms of anti-moralism, which insist on its coherence in different ways. The view I call ‘hard’ anti-moralism involves denying that those who are blameworthy are necessarily at fault, whereas its ‘soft’ alternative does not. I argue, first of all, that the unique problem of tragedy will remain intractable until we accept soft anti-moralism. I then explain why soft anti-moralism is only viable if we can account for a novel feeling of responsibility that is (a) of sufficiently ‘moral’ quality to do justice to the phenomenology of tragedy, whilst being (b) potentially fitting in the absence of fault.

This paves the way for Part Two of the dissertation, where I try to account for such a feeling in terms of ordinary agent-regrets in response to our failures to satisfy extraordinary reasons for action. Only once that is in place does it make sense to address the central question at the level of the collective, which I therefore do not attempt to do until Part Three.

CHAPTER ONE

Faces of Contingency in Our Lives

1.1. Contingency and Moral Luck

Life might not be such a fraught affair if how it is going did not matter so dearly to us. But for most of us, most of the time, it does so matter, and that is how things should be.¹ In recognition of this, we try to make our lives go as well as possible, doing what we can to navigate the world as it confronts us. That this does not always work out will be all too familiar. The world is a messy place, rife with contingency—it is seldom so obliging, and often downright recalcitrant. When contingency strikes, it frustrates whatever control we may have had over how our lives pan out. This raises the question: how should we respond to the workings of contingency in life?

Of the events beyond our control that impact our lives, some, we can agree, merely *happen to us*. Some of us are fortunate enough to inherit assets from our families, whilst others suffer the misfortune of having an earthquake destroy our homes. How, if anything, should we feel about this sort of thing? What, if anything, should be done? We may instinctively feel it somehow unfair that our lives could be made or ruined by such events, the occurrence of which, after all, lies beyond our control. We may then wish to

¹ See Chapter Three for elaboration on this claim.

neutralise the impact of such contingency, at least to some degree. There are important ethical views that try to explain why we should indeed respond in this way. For instance, luck egalitarians believe that social justice requires relations of true equality, but that truly equal relations only rule out those inequalities of goods that are undeserved—or down to ‘brute luck,’ as they say. Luck egalitarian justice thus enjoins us to neutralise the effects of contingency in our lives, for instance by setting up institutions that redistribute various goods from the beneficiaries to the victims of brute luck.²

Some believe there are life events beyond our control that do not merely happen to us, but rather, that we *play an active role in*. Not everyone shares in this belief, however, and it is easy to see why. We tend to think of our deeds as related to our powers of agency in some such way as to make us responsible for them, and the intuitive link between responsibility and control is very tight. This makes something like the ‘Control Principle’ highly plausible, the principle that we are not responsible for what we do unless it is in our control. I say ‘something like’ it because ‘responsibility’ and ‘control’ are said in many ways—and I will return to this in what follows. But we can, I think, see how accepting some such principle could easily bring responsibility and control so close together as to leave no conceptual room for the phenomenon at issue. In that case, life events beyond our control cannot be things we are responsible for doing, and to respond to them intelligibly, we will always have to take them as having merely happened to us.

Given all this, why would anyone believe that our responsibility for what we end up doing could be exposed to contingency? Those who share in this belief typically point to examples where the phenomenology seems to cut against the Control Principle. Such examples have been widely discussed since at least the seminal 1976 articles by Bernard Williams and Thomas Nagel, where they are introduced as cases of *moral luck*. The point of the examples is to show that it is perfectly ordinary for factors beyond our control to influence the feelings of responsibility we are apt to feel for what we end up doing. They achieve this by contrasting an agent who does one thing with a counterpart who does otherwise due to the lucky absence of the uncontrolled factor, and who differ in the responsibility they are apt to feel. Several types of moral luck are commonly acknowledged, which vary depending on the nature of the uncontrolled factor.

² For this view of the luck egalitarian ideal for redistributive institutions, see Richard Arneson (2011: 30).

‘Resultant luck’ depends on the results of conduct, as when two assassins shoot to kill, but only one actually kills, because the other has their bullet intercepted by the flight of a passing bird. ‘Circumstantial luck’ depends on the circumstances of agency, as when two officials are disposed to take a bribe, but only one actually does, because the other happens never to get offered a bribe. ‘Constitutive luck’ depends on how the agent is constituted, as when two bystanders might have intervened to stop a fight, but only one actually does, because the other was not brought up to be diplomatic and assertive.³ In each case, we are supposed to think that the unlucky agent will not only feel more responsible than the lucky one, but that they are *apt* to feel more responsible, even as we know that what they end up doing depends on factors beyond their control.

Although the phenomenology in moral luck cases is rather ordinary, it does not sit happily with the intuitive Control Principle. It sits so unhappily, in fact, that it appears to confront ethical thought with a dilemma. On the one hand, we could accept what the phenomenology delivers, namely, that it can be perfectly fitting for an unlucky agent to experience greater feelings of responsibility. But we can only do so by conceding that the responsibility we feel for what we end up doing can indeed, and fittingly, be affected by factors beyond our control, thereby denying the Control Principle.⁴ On the other hand, we could see it as unfair that the contingency of the world could impact our lives in this way, and thus insist on the Control Principle. But clearly, in so doing, we are forced to deny the ordinary phenomenology as illusory, leading to some form of revisionism about our everyday practices of responsibility, or at least our understanding of their grounds.⁵ Needless to say, neither horn of the dilemma is particularly appealing.

³ This is Nagel’s (1976) typology, minus what he calls ‘causal luck.’ I put causal luck to one side since I agree with Robert Hartman (2019: 3182) that it raises no philosophically interesting questions beyond those widely discussed in the free will debate, which is orthogonal to my concerns.

⁴ See John Greco (1995), Robert Hartman (2016) and Michael Moore (1997) for views along these lines.

⁵ Those who opt for this horn include David Enoch and Andrei Marmor (2007), Rik Peels (2015), Eduardo Rivera-López (2016) and Michael Zimmerman (2002). It should be noted that the revisionism that attends a denial of moral luck is especially troubling in light of existing legal practices. The mature criminal law in many jurisdictions supports liability for legal punishment on a pattern in keeping with the existence of at least some forms of moral luck.

In this chapter I explore the moral luck dilemma. My primary aim is to suggest that, amongst cases of moral luck, there is a particular subset that present a unique problem for ethical thought, the problem of how we should respond to *tragedy* in life. Resolving the problem of tragedy is the project of the dissertation as a whole. I begin in section 1.2 with some preliminaries. Section 1.3 explores some senses of ‘control’ as it may figure in the Control Principle, arriving at a reading capable of vindicating the phenomenology in a range of moral luck cases. With this in place, I go on in section 1.4 to discuss cases of tragedy. I explain the unique problem, namely, that even our refined Control Principle will not help us to escape the dilemma for tragic cases. Section 1.5 briefly concludes by outlining the shape of my preferred response, which I aim to substantiate in the chapters that follow.

1.2. Some Preliminaries

Before we begin, four preliminaries. First of all, in what follows, I limit my discussion to examples of resultant luck. Not only are these the cases that most interest me, but as we will see, the focus on these cases will allow considerations about both circumstantial and constitutive luck to arise in an organic way.⁶ For this reason, I think that proceeding in this way is not entirely arbitrary: it helps to display the conceptual relations between the canonical types of moral luck in a somewhat orderly fashion.

Secondly, it is customary, following H.L.A. Hart (2008: 211-30), to distinguish at least four senses in which we may be said to be ‘responsible.’ We may be: (1) ‘role-responsible’ for an action in that we are normatively expected to do it *qua* bearer of a role, (2) ‘causally responsible’ for an outcome in that we causally contribute to that outcome, (3) ‘liability-responsible’ for an action in that we are rightly subject to moral or legal consequences for it, and (4) ‘capacity-responsible’ in that we possess the requisite capacities for being responsible in the ‘role’ and ‘liability’ senses. As we saw, moral luck cases are supposed to show that we can be apt to feel responsible for what we did, even if it was beyond our control. Discussions of moral luck thus centre on the *liability* sense of ‘responsibility,’ that is, whether we are rightly subject to certain consequences for what

⁶ See subsection 1.3.3 for discussion of constitutive and circumstantial luck.

was beyond our control, specifically the *moral* consequences: feelings of responsibility. In this chapter I follow the standard practice in the literature of identifying these feelings with praise and blame.⁷ I limit my focus to cases of bad moral luck, where the aptness of blame, rather than praise, is at issue. My reasons for this are twofold: first, there are good philosophical reasons not to expect for the conditions of praise- and blameworthiness to be symmetrical,⁸ and second, my interest throughout the dissertation will be in the darker side of contingency, that is, where things are made to go awry.

When the propriety of feelings of responsibility is discussed in connection with moral luck, the tendency, as mentioned, is to identify those feelings with blame. However—and this is the third point—our blaming practices are normatively complex, and we must be careful not to be misled by that complexity. When we are blameworthy for what we did, we are apt to be blamed for it by first-, second- and third-persons: we can feel our own *remorse*, the *resentment* of our victim (if there is one), or the *indignation* of an onlooker. Remorse enjoys a certain priority here: when a person is blameworthy, the overall justifiability of second- or third-persons blaming them can depend on whether or not the first person shows remorse, whereas the reverse does not hold. For instance, if someone is rightly beating themselves up over what they did, that can make it seem inappropriate for the rest of us to pile it on, but if we aptly blame the person, that never seems to make inappropriate for them to blame themselves. Moreover, it may be that second- or third-persons lack the standing to blame a person who is in fact blameworthy, but that person, by contrast, cannot lack the standing to blame themselves. So, even when someone is blameworthy, the appropriateness of second- and third-person blame seems to depend on a whole lot more than that of first-person blame. To avoid confusion, then, it helps to think of ‘blameworthiness’ as the aptness of specifically first-person blame: of remorse as opposed to resentment or indignation.

The fourth and final point concerns the concept of ‘action’ we are working with. We can distinguish in ordinary thought between what we did and its outcomes, that is, what is brought about in or by doing what we did. According to G. H. von Wright (1963: 39-

⁷ See Hartman (2017) for a recent book-length treatment that proceeds on this assumption.

⁸ One widely acknowledged reason would be that, plausibly, praiseworthy action cannot be done in ignorance of its normative significance, whereas blameworthy action can (e.g. cases of negligence). Another more niche reason, owed to Susan Wolf (1980), would be that the kind of freedom required for praiseworthy action differs from that required for blameworthy action.

41), the 'in' and the 'by' loosely coincide with a finer distinction amongst the outcomes of an action, between what he calls its *results* and its *consequences*. A result of an action is an outcome that is constitutive of that very action, an ingredient of it, that is, one without which it cannot exist. By contrast, a consequence of an action is an outcome that is not an ingredient. Some examples: your dying is a consequence of my having shot you, but it is a result of my having killed you; my losing to you at chess is a consequence of your good endgame play, but it is a result of your having checkmated me; your being offended is a consequence of my having made a rude joke, but it is a result of my having offended you. When someone does something, we typically say its results are brought about *in* having done it, and its consequences *by* having done it.

Some philosophers conceive of actions in such a way that they never have their outcomes as ingredients. They believe our actions, strictly speaking, only ever have consequences—never results. Contemporary proponents of this view can be considered followers of Donald Davidson, who famously remarked that “We never do more than move our bodies: the rest is up to nature” (1971: 23). Notice that this view of action rules out the possibility of resultant luck from the off. When two assassins shoot to kill, but only one actually kills, there is, strictly speaking, nothing that the killer did that the would-be killer did not. After all, there is no having killed *X* without *X*'s having been killed—that outcome is an ingredient, a result, of having killed *X*—and so if actions never have results, then there can be no such action as having killed *X*. Thus, whatever the killer did, it can only be something from which the death of the victim is parsed out as a consequence, such as having shot to kill. Of course, that is no more and no less than what the would-be killer did. Sure enough, the killer's action also had the consequence that the victim died. On this view, however, *that* was up to nature, something that merely happened, for which the killer is no more responsible than are earthquake victims for losing their homes, or asset inheritors for gaining wealth. It would be apt, of course, for the killer to blame themselves for what they did do, which is having shot to kill. But no more so than the would-be killer, who, after all, did exactly the same.

In what follows, I assume the commonsense view that our actions can have results. I cannot do much to defend this assumption here, beyond gesturing towards a palpable

regress latent in the Davidsonian alternative.⁹ To see this, consider that even our moving our bodies, which is supposedly all we ever do, can be parsed into a piece of agency and its contingent outcome. Ordinarily, when we try to move our legs, our legs do as they are told. But it may be—a terrifying prospect—that we wake up one morning and, unbeknownst to us, we are paralysed from the waist down. We try to move our legs to get out of bed, only to find that we cannot. It turns out that moving our legs is an outcome of our trying, an outcome that, though normally certain, is here painfully revealed to be anything but. On the Davidsonian picture, this outcome has to be parsed out of our trying as a mere consequence. But again, even our trying can be parsed into agency plus consequence: we can decide to try to do something, and yet fail to even try. In order to locate the action, we will have to repeat the procedure. It is not obvious why we should expect this process to come to a principled resting place. In Nagel's metaphor (1976: 137), "the area of genuine agency [...] seems to shrink under this scrutiny to an extensionless point."

1.3. Senses of 'Control'

Cases of resultant luck purport to show that the results of our actions can influence whether or not we are blameworthy, even if they are in some sense beyond our control. They seem to suggest a tension between the ordinary phenomenology and the Control Principle, leaving us faced with a dilemma: either reject the phenomenology, or reject the Control Principle. So far, we have been operating with a rough and ready notion of the 'control' relevant to responsibility. In this section, I suggest that once we properly flesh out the notion of control, we see that the Control Principle is in fact consistent with a large swathe of the phenomenology in cases of resultant luck.

Why does control seem necessary for responsibility in the first place? It is, I believe, due to the strength of the following sort of case. Suppose grandma has made a beautiful birthday cake for Lex's younger sibling. Lex has been tasked with delivering it from the

⁹ For a defence of the assumption, see John Gardner (2018a: 64-71). The gestures that follow can also be found in Gardner (2018a: 59-60), though their basic pattern seems to be anticipated in the work of Michael Thompson (2008: 113-5).

kitchen to the front room, where his sibling is waiting with excitement. *En route*, however, a sudden gust of wind blows through the window, knocking the cake out of his hands and face down into the carpet. In so far as the gust of wind removes Lex's control over what he was doing, it is intuitive to say that he is not responsible for ruining the cake.

1.3.1. 'Control' as Reasons-Responsiveness

There are, however, clear cases where we are responsible for actions that were uncontrolled, in some sense, and precisely because they were uncontrolled. To vary the example, suppose that when delivering grandma's cake, Lex does so in a silly way that will impress his friends. He plays fast and loose with the cake, dancing about, hopping from foot to foot, pretending to nearly drop it, when—*splat*. He drops the cake. Intuitively, he is responsible for this, and yet he was acting in a way that was in a clear sense uncontrolled. Does this simple sort of case show that control is not necessary for responsibility after all?

Most of us, I expect, would think not. Pressed to say why, we would point to the fact that Lex's uncontrolled behaviour manifests an insensitivity (or at least under-sensitivity) to several important considerations—his grandma's efforts, the occasion, his sibling's feelings, and so on—that his behaviour really ought to have been attuned to. It ought to have been attuned to these considerations as an exercise of his power to be so attuned, a power to *respond to reasons*. So, we want to say: in acting in the uncontrolled way he did, Lex exercises a power to respond to reasons, albeit not very well. But it was, nonetheless, an exercise of that power—a power that Lex could have exercised better—and in *that* sense, he had 'control' over what he was doing, even as he did so in an erratic and 'uncontrolled' way. We arrive at the view that to be responsible for what we do, it needs to be under our 'control' in the sense of being an exercise of our power to respond to reasons, however erratic an exercise it may be.

I take it more needs to be said about control as reasons-responsiveness before it can be brought usefully to bear on cases of resultant luck. Our power to act responsively to reasons, we said, can be exercised better or worse. To say what counts as better or worse, we must first say what exactly this power is a power to do, and how it does it, when it does. I suggest the following: all being well, we act responsively to reasons when we *act*

in accordance with the balance of reasons, from the balance of reasons. So long as we act in accordance with the balance of reasons, from the balance of reasons, we exercise our power to respond to reasons aright. We do so, I claim, by being properly *receptive* to the balance of reasons and properly *reacting* on that basis.¹⁰ We are properly receptive to the balance of reasons when we are aware of the considerations bearing on our conduct in the context at hand, and how they stack up, so that there is some undefeated reason for action of which we are aware. We properly react to the balance of reasons when our awareness of an undefeated reason guides our conduct in practically competent accordance with it.

Note that the awareness I say is needed for guidance is not ‘awareness’ as ordinarily understood, that is, as an event in the stream of consciousness. We do, of course, sometimes respond on the basis of an explicit awareness of an undefeated reason. But our acting responsively to reasons need not be so intellectualised. For instance, if asked why they played a half-volley winner down the line, a tennis champion could correctly state a genuinely undefeated reason for which they did so. But it seems psychologically artificial to insist that, in the heat of the match, they did so on the basis of an explicit awareness of that reason. Their ‘awareness’ is a dispositional state, which can guide their behaviour without manifesting itself in the stream of consciousness.¹¹

I claim that to we exercise our power to act responsively to reasons aright is to be properly receptive to an undefeated reason and to properly react on that basis. We can expect, then, for poor exercises of this power to result from failures of receptivity or reactivity. In subsection 1.3.2, we will consider the ways such failures can happen, and the implications of this for the feelings of responsibility we are apt to experience. But to sensibly ask the question of whether a piece of conduct is a better or worse exercise in the first place, we should first have to ascertain whether or not it manifests the power *at all*. So, how can we know whether our conduct is guided in the right kind of way by awareness of (what we take to be) an undefeated reason?

¹⁰ The terminology is borrowed from John Martin Fischer and Mark Ravizza (1998), who put forward an early and systematic view of responsibility in terms of reasons-responsiveness. Though influenced by that important work, my own view, we shall see, deviates from their view in several ways.

¹¹ See Alexander Greenberg (2024: 356-7) for more on ‘awareness’ in this sense.

To know this, it is not enough to point out that our conduct satisfies (what we take to be) an undefeated reason. For instance, suppose that Molly, who has nothing better to do, goes on TikTok and watches some genuinely informative content about climate change. We may be tempted to think that, in so doing, she acts responsively to reasons. But now suppose that she would have gone on TikTok no matter how cranky the content she watches (e.g. conspiracy theories; racist propaganda), or—no matter the content—if she were to find herself in a burning building. That would, I think, show that her watching climate content on TikTok is not reasons-responsive after all. For it now seems that she just *happens* to satisfy an undefeated reason: since she would have done no differently had (what she took to be) the reasons or their balance been different, what she actually did does not seem rationally guided in the right kind of way—it does not seem to manifest her power to respond to reasons at all.

So, the test for the reasons-responsiveness of what we actually did is whether we would have responded to a range of counterfactual reasons were they present.¹² The number and importance of reasons in that range may vary, and with it, I believe, the degree of reasons-responsiveness of what we actually did. Since our ‘control’ of what we do in this sense comes in degrees, it is natural, and I think correct, to suppose that our responsibility does too. There is probably no principled way to say where the threshold lies, but we can, it seems, discern conduct that clearly falls below it: in profound cases of addiction, mental illness or cognitive disability, for example. I will say, in such cases, that people are *exempt* from responsibility for their conduct.¹³

¹² To be clear: the reference is to other things we *would* have done had there been other reasons, the point of which is to get an empirical handle on whether our actual conduct was reasons-responsive. It is not to be confused with any sort of reference to other things we *could* have done, with reason or otherwise, taken as a metaphysical condition on the ‘freedom’ of our actual conduct. The view presented here is meant to be a compatibilist view. See Michael McKenna (2022: 29-32) for helpful discussion of this point.

¹³ See Gardner (2003: 159-61) for a powerful illustration of exemption, as contrasted with excuse, as a way to deny ‘responsibility.’ Gary Watson (1987: 260-1) explains exemption in a way that fits the present mould, though see Chapter Three for deeper problems with neo-Strawsonian proposals like his.

1.3.2. Failures of Receptivity and Reactivity

Once we know of a piece of conduct that it is an exercise of a power to act responsively to reasons, we can sensibly ask whether it was aright as an exercise of that power. Of course, the question does not usually arise unless the conduct has obviously bad results, results that make the conduct a *prima facie* failure to satisfy undefeated reasons, making us presumptively blameworthy. Why does this make us presumptively blameworthy? I suggest it is because, as responders to reasons, it should be of first importance to us all that we do not exercise our power to do so poorly. And if it looks like we failed to satisfy undefeated reasons, that is precisely what it looks like we did. That *blame* is what seems apt is why it is usually thought that the undefeated reasons against our conduct must be the especially important ones we call ‘duties’ or ‘obligations.’ I am not entirely sure about this assumption, so they do not figure in every example in what follows. I hope that the intuitiveness of the examples helps to bring out why I hold this view, but if it does not, then the reader should feel free to substitute the undefeated reasons in the examples with duties or obligations.¹⁴

When we are presumptively blameworthy, then, there is some presumptive way that we have failed to exercise our power to act responsively to reasons aright. This can result from failures of receptivity or reactivity. Take failures of receptivity first. We are not properly receptive when we lack awareness of the offending undefeated reasons against our conduct, or of the fact that they are undefeated, including reasons that are facts about risks (or more simply: ‘risks’). In that case, we either fail to react to reasons at all, or we react on the basis of whatever reasons we take to be undefeated in the circumstances at hand. Either way, we fail to properly react to the balance of reasons as they confront us: we conduct ourselves with a poor receptivity to the undefeated reasons, and that conduct, other things equal, will fail to satisfy those reasons.

¹⁴ To reiterate: I am unsure that, to be blameworthy, we must have done anything especially bad, over and above its not being in accordance with and from the balance of reasons. There is a perfectly ordinary sense of ‘blame’ which we could use to describe kicking ourselves for relatively minor mistakes for which we are culpable. I believe that where we are apt to kick ourselves for these mistakes—not a hard kick, but a kick nonetheless—they manifest a failure of reasons-responsiveness that meets the conditions for blameworthiness described in this section. So, I do not see why we should not say that the kick expresses ‘blame.’ For more on this topic, see Chapter Two.

Standard cases of *negligence* involve people who conduct themselves with such a pattern of awareness. For instance, imagine a father who drives to school to collect his son because it is the hottest day of the year.¹⁵ Upon arriving, he is greeted by the school principal who calls him in to discuss an incident involving his son, which, due to its seriousness, ends up being a long meeting. Several hours later, father and son walk back to the carpark in the brutal heat. To their horror, they find the family dog left airless in the back seat of the car, unconscious from heat exhaustion. The dog has clearly slipped the father's mind; and with it, whatever reasons there were—risks or otherwise—for him to do what would have kept the dog safe on such a hot day: reasons we can assume were undefeated in the balance. *Prima facie*, the dog's heat exhaustion is the result of the father's negligent conduct.

Let us now turn to failures of reactivity. Failures of reactivity presuppose proper receptivity in the above sense, and so cannot be negligent. Sometimes we are aware of an undefeated reason, which may be an undefeated risk, but we deliberately act in ways that go against that awareness. Though we know there is an undefeated reason not to do as we do, we go ahead and do it anyway. Our conduct bespeaks a knowing disregard for whatever gives us the undefeated reason. We can start by considering cases where that knowing disregard is purely instrumental: we knowingly disregard whatever gives us the undefeated reason, possibly an undefeated risk, in order to do what we are aware is defeated in the balance.

Such failures of reactivity explain standard cases of what we would call *knowing* or *reckless* conduct. To vary the previous example, suppose that the father, when called in by the principal, was aware that he left the dog in the car, possibly without air (he does not recall), and that, either way, if the meeting happens to drag on, there will come a point at which the dog is seriously imperilled by the heat (though exactly when he is not sure). In that case, he is clearly aware that he is risking the dog's safety by going to the meeting (at least without first attending to the dog)—a risk that tips the balance of reasons against doing so—and yet he goes anyway, wanting to find out immediately what happened with his son. To that extent, he is reckless in so doing. And if, at some point in the meeting, he becomes quite certain that the dog will perish unless he attends to it right away, but continues with the meeting anyway, then his recklessness as regards the

¹⁵ The example is a variation on one originally given by George Sher (2006: 286-7).

dog's safety tips over into knowingly letting it suffer harm. Either way, unlike the case where the father is negligent, it is not his receptivity that is the problem: and he seems to be all the more culpable for that reason.¹⁶ His pattern of awareness puts him in a better position to properly react; thus his failure to do so is a *prima facie* faultier exercise of his reasons-responsiveness.

In other cases, our disregard for what gives us the undefeated reason is the very point of our conduct. We are aware that the balance of reasons speaks against a course of conduct, but we do it anyway: not in order to do something else, but precisely to effect the bad result. We standardly describe knowing conduct of this kind as *malicious*. To return to our example, suppose that the father had long considered the family dog something of a nuisance: the dog is loud, disobedient, aggressive, and not to mention resistant to housetraining—with all the cleaning fees that entails. He had often thought about how he could be rid of the dog, without sparking conflict in the family. So when called in by the principal, he spots his chance to do just that. He is aware of his undefeated reason against going into the meeting, but he does so anyway, finding ways to extend the meeting as long as he can: not out of any curiosity about his son's incident, but to ensure that the family dog perishes. His knowing disregard for the dog's safety is the very point of his conduct, which is why it strikes us as malicious.

¹⁶ In this, I demur from Seana Shiffrin's way of distinguishing knowing (including reckless and malicious) from negligent conduct. Her examples (2017: 202-3) of paradigmatic 'negligence,' are, to my mind, either under-described (so that they need not be negligent) or more properly described as 'reckless.' I do not deny that, ordinarily, we might speak of her preferred cases in terms of 'neglect,' but here I would be wary of ordinary language. For example, we may ordinarily describe someone who maliciously lets their partner's plants die whilst they are away (following an argument, say), as maliciously 'neglecting' the plants. That is no reason to suppose our technical understanding of negligence must accommodate such a pattern of awareness. We need to draw the lines between paradigm cases somewhere, but Shiffrin, I think, does not. About this, she is quite explicit (2017: 205): "The merely negligent agent may tolerate knowledge that a small portion of her duty will go unsatisfied, whereas the reckless agent may tolerate knowledge that the bulk or the whole of her duty will go by the wayside. The reckless agent, then, represents the extreme or limit case of negligence." Both uses of 'may' here are indicative: our ordinary understanding of 'negligent' agents *may* and so *may not* tolerate knowledge (i.e. in many cases they are quite ignorant), whilst our ordinary understanding of 'reckless' agents *must* tolerate knowledge. I therefore do not see what is lost in taking ignorance as paradigmatic of negligence, besides accommodating ordinary linguistic usage, which, as we have seen, may not be a helpful guide to the phenomena.

One final way that that our reactivity can fail does not involve knowing or reckless disregard of what gives us the undefeated reason at all. We are properly receptive to the undefeated reason and we do not deliberately act in ways that go against that awareness, yet we still fail to properly react. Although our conduct is guided by the undefeated reason, it is *practically incompetent*, so that, other things equal, we fail to satisfy the reason we are guided by. It can be hard to imagine such failures where the conduct in question is less obviously skilful. But even the most mundane of our activities, I believe, involve practical competencies—and they can be hard to see precisely because their smooth functioning is so pervasive.¹⁷ Suppose that the father follows the principal as far as the school entrance, but then, remembering the dog, excuses himself for a moment to leave the dog somewhere safer. But he somehow gets lost for half an hour on the way back to the carpark, and then, when he finds it, cannot remember where he parked the car. After forty-five minutes of unsystematic and inefficient searching, he finally locates the car, only to find that he has now misplaced his keys. After another half hour of rummaging around his bag, he eventually checks his pocket, where the keys are located. He unlocks the car and lo and behold, the dog has fallen unconscious with heat exhaustion. The father in this case is not unaware of the dog, nor does he not try hard to keep the dog safe, as the balance of reasons would have him do. Rather, he does not try well enough, competently enough—and as a result, the dog has come to harm.¹⁸

¹⁷ Cases of practical incompetence are often not implausibly reimagined as failures of receptivity, so it can be tempting to suppose that this category does not exist. I admit that the line between them is sometimes blurry, but to my mind, it is psychologically artificial to suppose that those who are fumlbers, bunglers, clumsy, klutzes, bottlers, and so on, are always so because they lack awareness of a reason, or indeed propositional knowledge of any kind.

¹⁸ Often, in cases like this, it is a live question whether the conduct even passes the test for reasons-responsiveness at all. Admittedly, here, the father's incompetence is so spectacular that it seems unlikely that he is not pathological in a way that exempts him from responsibility *tout court*. But nevertheless, whether or not that is the case is an empirical matter: if he would have responded differently to a range of possible reasons, then his actual conduct is reasons-responsive—however bizarrely. For instance, there are probably ways we could extend the case such that it is plausible that the father would not have been so incompetent had he left his baby, rather than the family dog, in the car.

1.3.3. Justifications and Excuses

In each scenario above, the father's conduct results in harm to the family dog, so it appears *prima facie* that he acts contrary to the balance of reasons. This makes it look like he failed to exercise his reasons-responsive powers aright, which is why, we said, he is presumptively blameworthy. So it makes sense, then, that he can get himself off the hook by showing that, though the way he exercised his powers did indeed have bad results, he either: (a) nonetheless did so aright given the circumstances, or (b) did not do so aright, exactly, but did so as well as could be expected of him given the circumstances. If there is some true story to tell that makes either case, his conduct does not in fact reflect badly on him as a responder to reasons. It would show, that is, that he is not at fault in the way he exercised his power to respond to reasons.¹⁹

Stories of type (a) are *justifications*. To act with justification is to act in accordance with, and from, the balance of reasons. Thus when our action is justified, it is as it should be as an exercise of the power to act responsively to reasons. A true justification shows that, though we did indeed deliberately bring about or run the risk of a bad result—something there was indeed *prima facie* undefeated reason to avoid—the circumstances were such that, in fact, this was the only way we could satisfy the balance of reasons, which is why we acted as we did. It shows, that is, that we deliberately went against the *prima facie* undefeated reason in order to satisfy a reason that was *bona fide* undefeated in the circumstances. Imagine, for instance, that *en route* to the school, the dog had gone mad and repeatedly mauled the father half to death, leaving him no other option but to lock the dog in the car in an act of proportionate self-defence. Or perhaps, on arriving at the school, the father—who works as a government agent—is correctly tipped off that the school is about to be subject to a chemical attack, making it a necessity to rush immediately to the principal's office, which is the only place it will be safe for the next few hours. Or, again, perhaps the father acted under duress of a terrorist's credible (if

¹⁹ What follows is meant to more or less reproduce Gardner's rich body of work on this topic. For his views on the relationship between reasons-responsiveness, justifications and excuses, and responsibility, see his (2003: 157-9; 161). For the view that excuses appeal to the strength of justified beliefs or feelings, see (2007a: 109-11), and for the view that the relevant standards of justification involve reference to what can be reasonably expected of an occupier of the relevant role, see (2007b: 128-31). Finally, for the view that we are at fault for what we did when we are neither justified nor excused in having done it, see (2005: 110-5).

odd) threat to kill his whole family unless he attends the meeting for several hours. In any of these scenarios, the father could be justified in knowingly leaving the dog in the car to die, or to run the risk that this will happen.

Stories of type (b) are *excuses*. Unlike justifications, excuses concede that the bad results did indeed result from a failure of reasons-responsiveness, that is, that there was indeed undefeated reason not to do as we did. They go on to say, however, that a closer look at the circumstances reveals that, though we failed to satisfy an undefeated reason, we were nonetheless as responsive to reasons as could reasonably be expected of us. To that extent, a true excusatory story shows that our failure to act properly responsively to reasons was not our fault. In general, the way they try to do this is to point to features of the circumstances that amount to undefeated reasons to believe or feel a certain way, saying that, though our conduct was ultimately unjustified, we only acted as we did on the strength of beliefs or feelings that were themselves justified. The appeal to justified belief applies more easily to failures of receptivity, as failures of reactivity presuppose that we are properly receptive to reasons, such that it should not be news to point out that our pattern of awareness (and so belief) is epistemically justified. By contrast, an appeal to justified feelings could apply to failures of either kind.

Examples of excuses that appeal to justified feelings include provocation and some cases of duress. Suppose that the father, upon being called in by the principal, deliberately leaves the dog airless in the car, knowing that the meeting will (or will be likely to) go on for so long that the dog will perish. But suppose also that the dog is highly beloved by the father's spouse, who has, over many years, been serially abusive in the most depraved ways, making life a living nightmare for the whole family, and threatening violence if he ever reports on the abuse to the authorities. The father is by now, and understandably, an emotional wreck, living in a constant state of fear, insecurity and bubbling resentment. Just before the school run, the spouse had, for the fifth time this week, had another violent outburst. It was the last straw for the father, who, in a flash of rage *en route* to the school, got it into his head that he would exact vengeance by hurting the dog he knew his spouse held so dear. In so doing, he acts unjustifiably: to exact revenge, however sweet, is arguably never a reason to do anything, let alone a *bona fide* undefeated reason to deliberately hurt the dog. Nonetheless, he appears to act in this unjustifiable way on the strength of feelings—rage, resentment—that are themselves quite plausibly justified, given the long campaign of terror and abuse

he has suffered at the hands of his spouse. The father's actions, though unjustified, are quite plausibly excused as a reaction to his spouse's provocations.

Or we could return to the scenario where the father acts under duress. Only this time, what has been threatened is not something the reason to avoid which in fact defeats his reason not to do what will (likely) harm the dog. Suppose, for example, that the threat is to spread social media rumours that could destroy his reputation, unless he attends the meeting for several hours. Here, if he acts to avoid what is threatened, thus (running the risk of) harming the dog, he will not be justified in so doing. So, in the event that he does, then though it is true that he acts under duress, the relevance of that circumstance to the evaluation of his conduct differs from the duress case we considered before, where the terrorist threat to his family in fact justifies his acting to avoid it. Nevertheless, it may well be that what leads him to attend the meeting is a genuine fear of the threat being realised, a fear of reputational damage that is itself quite justified. In that case, the relevance of duress is not that it gave him an undefeated reason to do what will (likely) harm the dog, but rather an undefeated reason to fear the repercussions of not doing so, on the strength of which he then acted. The father's actions, though unjustified, are again plausibly excused as a reaction to the circumstance of duress.

Duress may also excuse in cases of incompetence. Recall, for example, the scenario in which the father actually tries to secure the dog's safety, but fails to do so out of *practical* incompetence: he somehow really struggles to find the car park, the car, and finally the keys. Suppose, this time, that the principal—a lover of dogs, and an unscrupulous social media power-player—threatens to spread viral takedowns of the father unless he attends to the dog, which the father was going to do anyway. And suppose that the fear of reputational damage was so overwhelming that it led to the practical incompetence with which he tried to attend to the dog. In that case, again, he is unjustified in his failure to save the dog, but quite possibly excused. It should be noted, however, that the incompetence need not be practical for a duress excuse to apply. To see this, consider that negligent conduct is *epistemically* incompetent. Return to the case in which the dog simply slips the father's mind. But suppose, this time, that upon arrival at the school, the principal issues a threat to the effect that unless the father steps into the office immediately to discuss his son's behaviour, they will take to social media to spread viral takedowns of his poor parenting. And suppose that this instilled in the father a fear so overwhelming that it led him not to think clearly, as he usually would,

which is what led him to forget all about the poor dog. In this case—as in the case of a failure of reactivity—though the father’s unjustified conduct results in harm to the dog, the failure of receptivity may be excused as owed to a justified fear of the principal’s threat.

Although negligence is always a failure of receptivity, the lack of awareness it involves need not be caused by the strength of any feeling. Sometimes, when we are not aware of an undefeated reason, or of the fact that it is undefeated, our ignorance is due to the strength of certain beliefs that are explainable by features of the circumstances without the mediation of any feelings. This brings us to excuses that basically appeal to beliefs as opposed to feelings. For when those beliefs are not only explained but justified by circumstantial features, the resulting ignorance intuitively excuses our conduct, showing our conduct to have been non-negligent.

Suppose that, as in the original case, the father is indeed unaware of his undefeated reason not to attend the meeting, namely, that if he does so, the dog will collapse from heat exhaustion. Only this time, his car is fitted with a tried and tested app, which allows him to turn on the air-conditioning remotely with his phone, and which live-streams the inside of his car. He remotely turns on the air-conditioning as he enters the principal’s office, consulting his phone throughout the meeting to check on the dog. The live-stream seems to display that the dog is alive and well. Little does he know, however, that the app is buggy: the air-conditioning is not on, and what he sees is in fact yesterday’s footage. So when father and son return to the vehicle, they find, to their horror, that the dog has collapsed. All along, the father was ignorant of his undefeated reason to attend to the dog, a reason that in fact made staying in the meeting unjustified. But his ignorance stems from a belief that the dog was alive and well, a belief he was perfectly justified in holding, given what the tried and tested app displayed. So, though the father unjustifiably harms the dog, the lack of awareness from which he does so is plausibly excused by an ignorance that was quite justified under the circumstances.

Excusatory stories point to features of the circumstances to which we plausibly responded justifiably in belief or in feeling, which in turn guided the conduct at issue, thus making the case that we acted as reasons-responsively as could be expected of us given the circumstances. Why is pointing to such justified beliefs or feelings necessary for making that sort of case? It is because whether we were justified to respond with the beliefs or feelings we did—whether in content or in degree—very much depends on the

standards that can be reasonably expected of us given the capacity in which we acted.²⁰ Suppose we assume that the father acts in his capacity as a dog-owner. We may then wonder if the father, *qua* dog-owner, could perhaps have been expected to show more resilience in the face of threats to his social media image than he did, in the relevant duress cases. If we say ‘yes,’ we are saying that the ignorance or fear from which the father acted was, though intelligible, not after all justified, and thus that the various circumstances of duress are not enough to excuse him. It seems harder to say ‘yes’ to an expectation of greater resilience in the face of threats to his family, or in the face of a long campaign of serious domestic abuse. But now suppose the father acts in his capacity as an elite government agent, and that the dog’s safety is essential to protecting national security. In that case it seems much more certain that, ‘yes,’ he should have been more resilient to the threat to his social media image, and much more arguable that, ‘yes,’ he should have shown greater resilience in the face of a threat to his family, or a long campaign of domestic abuse. Whether the circumstances are sufficient to excuse thus very much depends on the capacity in which we acted.

Whether as a dog-owner or a government agent, the answer ‘yes’ may invite the complaint that it is unfair to expect more resilience from the father, since he did, as a matter of fact, show as much resilience as he was capable of. He was simply not brought up to be especially resilient, and so incapable of showing the resilience expected of him. Perhaps a better raised dog-owner or government agent could indeed have shown more resilience under the same circumstances, but it is surely unfair to let such morally arbitrary factors like upbringing determine whether he is on or off the blame hook—or so the complaint goes. The complaint, in other words, is that the father is subject to something like *constitutive* luck. Though a familiar complaint, the problem with it is that it misunderstands how excuses work to exculpate. Excuses show that although we failed to act properly responsively to reasons, we nonetheless responded to reasons as well as could be expected of us given the circumstances, so that we cannot be faulted for our

²⁰ Though, as we saw, the reference to role-relativised expectations is inspired by Gardner, it dovetails nicely with what Dana Nelkin and Manuel Vargas (2024: 48-51) have recently called their *ecological* version of (agent-based) reasons-responsive theory. They say that “a satisfactory theory of culpability must be sensitive to the context,” and that “the context matters for what we can reasonably expect of those agents.” The ‘context,’ in turn, may include “socially and normatively significant features,” thus making “our norms of [culpability] assessment are sensitive to roles, interests and expectations.”

failure. So, if there are standards of belief or feeling that can be reasonably expected of us, it makes no difference if we fail to live up to them because we are constitutionally incapable of doing so. All that matters is that we fail to live up to them. Thus, it reflects no less badly on us as responders to reasons if we are incapable of doing so: indeed, if we are constitutionally unfit for our role—like the father in any ‘yes’ case where he is genuinely incapable of greater resilience—then that is precisely what reflects so badly on us.

Even conceding this point, we may still detect something morally arbitrary about answering ‘yes’ and blocking the father’s excuse. For even if the father fails to show the resilience expected of him in virtue of his role (whether he was able to or not), we may point out that he is no more insufficiently resilient than a similarly insufficiently resilient person who happens to be lucky enough not to be confronted with a comparable opportunity to be insufficiently resilient. We may, that is, complain that the present view unfairly subjects the father to *circumstantial* luck. But what exactly does it mean to say we have been comparatively unlucky to find ourselves with an opportunity to act insufficiently virtuously, or, by the same token, without an opportunity to act virtuously? The point is easier to see with the latter sort of case, so let me take that as an example. If our complaint is that we have been unlucky not to find ourselves with an *identical* opportunity to our comparator, whether numerically or normatively, then that seems a rather childish thing to fret about.²¹ More likely our complaint is that we do not enjoy so *great* an opportunity as our comparator. For instance, we all get to exhibit courage sometimes, but a small minority of us get to do so in high political office in actions with world-historical consequences. This is perhaps a more serious complaint, but it is, I believe, defused by the plausible thought that an equivalent greatness of virtue (and of vice) can be manifest in action in lofty and mundane contexts alike. I cannot do much to defend this claim here, besides the suggestion that some such lesson can be discerned in Hannah Arendt’s (2022) commentary on the ‘banality of evil’ in fascist Germany, where even the most everyday acts of humanity (and inhumanity) could be seen in their enormity to ethical life.

²¹ In saying this, I mean to agree with Enoch and Marmor (2007: 425).

1.3.4. 'Control' as Faulty Reasons-Responsiveness

The preceding discussion may seem to have taken us very far from the Control Principle. But in considering the ways our 'control' in the sense of a power to act responsively to reasons can go awry in its exercises, and the ways we absolve ourselves of the blame we presumptively attract, we have in fact been working our way towards the conditions of feeling 'responsible' in the sense of 'blameworthy,' which speaks to the Control Principle as commonly understood. To begin to join the dots, let us return to the case of Lex.

In the original case, Lex is delivering grandma's birthday cake to his sibling, when a gust of wind blows in through the window and knocks the cake out of his hands. We then considered a variant where Lex once again drops the cake, but this time because he was messing around in a bid to impress his friends. In both cases, we said, Lex's actions are in some sense 'uncontrolled.' But now, again in both cases, we can say that Lex was in 'control' of his actions in the sense that his delivering the cake was an exercise of his power to act responsively to reasons. We may, then, wonder why Lex is intuitively *not* blameworthy in the first case, but *is* in the second. In terms of the Control Principle: in what sense does Lex have 'control' in the second case, such that he lacks it in the first?

Given what we said in the last two subsections, I believe we have the resources to provide a good answer. In either case, when Lex drops the cake, he goes awry in exercising his power to act responsively to reasons. But in the first case, Lex can easily rebuff any presumption of blame for what he did by pointing to circumstantial features that will get him off the hook. As far as we know, he has no justification available: there was no reason, let alone an undefeated reason, for him to drop the cake. He does, however, almost certainly have an excuse. The shape it takes will depend on what kind of failure of reasons-responsiveness we attribute to him, but given what we know, the only charge that can sensibly be made is negligence: we claim he lacked awareness of an undefeated reason not to deliver the cake without first taking precautionary measures, e.g. checking the forecast, closing the window, and so on. But under normal circumstances, we have no reason to take such measures before doing something so simple, and Lex is, we can assume, perfectly justified in believing that the circumstances were normal—however abnormal they in fact were, owing to the sudden gust of wind. So, although he lacks awareness of the undefeated reason, he can plausibly claim that this ignorance stemmed from a false but justified belief held in response to the

circumstances, which turned out to be less normal than they seemed. In the first case, then, Lex is excused: though he dropped the cake, he was not at fault in so doing.

In the second case, however, Lex *does* seem to be at fault. Given what we know, this may be due to negligence: a lack of awareness of the risks associated with his conduct. It is not obvious how he could get himself off this particular hook. Normally, if we are undertaking physical tasks, an erratic attempt at physical comedy runs the risk of undermining our success in those tasks. The circumstances were normal in this respect, and Lex has no reason to believe otherwise, let alone an undefeated reason. It could, of course, be argued that Lex was ignorant of this principle. But past a certain age, which we can assume Lex has reached, we can be reasonably expected to know such things, and so his ignorance remains unjustified. Or we might appeal to Lex's feelings: perhaps he got so excited by the occasion, or the prospect of pleasing his friends, that he forgot all about the risks? But even if this is what happened, that degree of excitement is not plausibly justified: again, past a certain age, he can be expected not to be so easily excitable as to become totally oblivious to certain risks.

More likely, though, Lex is at fault in being aware of the risk but running it anyway: not out of malice, but sheer recklessness. This time he could try to justify himself: he could, for instance, claim that pleasing his friends was genuinely more important than not dropping the cake, and that he could not achieve this higher purpose without running the risk of dropping the cake. Of course, his parents would no doubt be quite unimpressed by such a story. Alternatively, he could try to excuse himself. But here, as everywhere else, it is hard to see what he could plausibly say—or not, at least, without adding details to the case. He could try to appeal to some feeling that led him to mess around as he did, thus running the risk of dropping the cake. Such an appeal may or may not be made in tandem with the claim that his messing around was, for all its appearance as skilful agency, in fact practically incompetent. Either way, it makes no difference. For, given what we know, there are no circumstantial features present that would plausibly justify whatever feeling he appeals to, as would be necessary for such an excusatory story. There is simply nothing in the circumstances that plausibly provoked or threatened him, no basis for making an excuse of provocation or duress.

We can summarise as follows. In both cases, Lex fails to properly exercise his power to act responsively to reasons, but only in the second case does the way he fails reflect fault on his part as a responder to reasons, for only there is he unable to point to

abnormal features of the circumstances that responding to which led him astray. In the first case, by contrast, it was abnormal circumstances that led Lex astray, so that, in going astray, he was without fault as a responder to reasons. I believe we can state the difference between them in terms of a perfectly ordinary sense of ‘control.’ In the first case, it is the abnormality of Lex’s circumstances (viz. the gust of wind) that renders him quite faultless, pushing what he did in that sense beyond his control. In the second case, no feature of the circumstances renders him faultless; his failure was his own fault, and was in that sense—the same sense—perfectly within his control. ‘Control,’ we can see, is said in many ways. In both cases, Lex had ‘control’ of his actions in that he was exercising his power to act responsively to reasons, and those exercises can both be described as ‘uncontrolled.’ But what separates them is this. In the first case, dropping the cake was beyond Lex’s ‘control’ in the technical sense that he was faultless in so doing, whereas in the second case, it was within his ‘control’ in the same sense: he was at fault. *This* is the sense of ‘control’ that figures in the Control Principle on ‘responsibility’ in the sense we are interested in, that is, as blameworthiness.

Perhaps a more perspicuous name for our refined Control Principle would be the ‘Fault Principle,’ which I state as follows.²²

Fault Principle: We cannot be blameworthy for action that is not a faulty failure to properly exercise our power to act responsively to reasons.

With this refinement in hand, I believe we are in a position to see how a proper understanding of the Control Principle—the Fault Principle—is in fact consistent with the phenomenology in many standard cases of resultant luck. To see this, take the assassins example: two assassins shoot to kill, but only one actually kills, as the other is lucky enough to have their bullet intercepted by a passing bird. We sense that it would be more fitting for the unlucky assassin to feel responsible, even though we know, as the lucky counterpart brings out, that what they end up doing depends on factors beyond their control. The intuition behind the Control Principle tells us it is somehow unfair for what is beyond our control to affect our responsibility for what we end up doing, thus imploring us to deny the phenomenology: if it is fitting for either assassin to feel

²² We consider objections to the Fault Principle in Chapter Two.

responsible, it is fitting for them both. On the Fault Principle, however, both assassins had ‘control’ over what they did in the sense that they were fault in so doing, but the difference in what they did explains their differential blameworthiness. Both assassins, we can assume, had an undefeated reason not to shoot to kill, which they were both at fault in failing to satisfy. But both, we can also assume, also had an undefeated reason not to kill, which only one of them—the unlucky one—is at fault in failing to satisfy. However faulty the lucky one may have been for shooting to kill, they cannot have been fault for killing, because they did not kill. The unlucky assassin is at fault for one more failure in their responsiveness to reasons, which makes them, by the Fault Principle, blameworthy for one more failure, and a graver one at that.

The assassins case involves malicious (or at least knowing) conduct, but the phenomenology in standard cases of resultant luck involving reckless or negligent conduct can be accommodated on the same pattern. This is because if an action does not fail to satisfy a *prima facie* undefeated reason, it is not a *prima facie* failure to properly exercise our power to act responsively to reasons, and so the Fault Principle rules out that the lucky agent is *prima facie* blameworthy for the graver failure. So, if two revellers drink-drive home from the pub, but only one actually injures someone in a car accident, then the Fault Principle supports the intuition that the injurer is more to blame than their lucky counterpart, because they are to blame not just for recklessly drink-driving, but for recklessly injuring someone in so doing. Or if two drivers are ignorant of the risks obviously indicated by the loud grinding noise emanating from their vehicles, but only one actually injures someone when their brakes fail, the Fault Principle supports the phenomenology in the same way. The injurer is more to blame than their lucky counterpart, as they are to blame not just for negligently setting out in a dodgy vehicle, but for negligently injuring someone in so doing.

1.4. Tragedy and the Fault Principle

Contrary to initial appearances, insisting on the Control Principle need not be so at odds with the ordinary phenomenology in cases of resultant luck. The Fault Principle is an independently plausible way of taking the Control Principle that resolves the dilemma, one the acceptance of which does not force us to revise our everyday practices of

attributing responsibility. Or not, at least, for those cases of resultant luck where both the lucky and unlucky agents are intuitively at fault.

But not all cases of resultant luck involve agents who are at fault. In another range of cases, both agents intuitively act *faultlessly*. The way they conducted themselves in doing what they did does not reflect badly on them as responders to reasons because they were, if not justified, then at least adequately excused. That is, when we cast a clear eye on their circumstances, we see that they exercised their power to act responsively to reasons aright, or, if not quite aright, then at least as well as could be reasonably expected of them. Yet for the unlucky one, contingency strikes, so that their faultless agency has disastrous results—results that, of course, do not descend upon their lucky counterpart. As cases of moral luck, the phenomenology is meant to attest to the aptness of a feeling of responsibility for these disastrous results, despite our knowledge that they were not the agent's fault: it was beyond their control.

We can illustrate the ordinariness of the phenomenology by considering the fate of familiar heroes from Greek *tragedy*. Take Sophocles's tragic hero Oedipus for example. Oedipus is known for unwittingly fulfilling the prophecy of Tiresias that he would one day kill his own father and marry his own mother. In the course of the play's events, it is revealed that the man he killed at the crossroads near Delphi was in fact his father, the Theban king Laius, and that the woman he married was in fact Laius's widow, his own mother, the queen Jocasta. Upon learning the bitter truth, Jocasta hangs herself, and Oedipus, over Jocasta's lifeless body, gouges out his own eyes in a fit of despair. The contingency of the world wrong-foots our characters at every turn, making for a tragic affair of quite epic proportions. For our purposes, the point would be that Oedipus's extreme reaction to what he turns out to have done (*parricide, incest*) strikes us as the expression of a feeling that is not only intelligible, but genuinely apt. And this is so, in Martha Nussbaum's (1994: 43) view, despite the facts of Oedipus being presented such that "So far as the intentional content of his desire is concerned, Jocasta is simply a well-placed eligible stranger," that "his aggressive action against Laios is in and of itself culturally acceptable, a counterattack in self-defense," and moreover, "Nor is there any sign that Oedipus has at any level hidden knowledge about the identity of the stranger he kills." Nussbaum's point, in short, is that Oedipus's parricide and incest are, though perhaps unjustified, surely at least excused in the circumstances at hand.

Another example would be the Mycenaean king Agamemnon as portrayed by Aeschylus. The events of the play surround Agamemnon's homecoming following his defeat of the Trojans, a victory that was made possible, famously, only by sacrificing the life of his innocent daughter Iphigenia. This is, of course, a brutal thing to have done, but it was, in Aeschylus's telling, an act necessitated by the gods. The expedition against Troy had been commanded by Zeus, and an angry Artemis had, moreover, contrived to make the expedition's departure impossible without the sacrifice of Iphigenia. With adverse conditions putting the soldiers in peril, inaction would not only have violated divine obligation, but would likely have caused everyone involved to perish. In such extreme circumstances, it may well appear that the sacrifice of Iphigenia was the lesser of two evils: Agamemnon is portrayed as doing what he had to do. Nonetheless, it may seem that he goes awry in so easily reconciling himself, in feeling, to his act. This impression, as Nussbaum observes (2001: 33), is shared by Aeschylus's Chorus: "The sacrifice of Iphigenia is regarded by the Chorus as necessary; but they also blame Agamemnon." Modern readers take Agamemnon to be apt to feel responsible for what he did, even as we perceive, with the Chorus, that this was justified in the circumstances, rendering him faultless in so doing.

"Greek tragedy," in Nussbaum's analysis (2001: 25), not only "shows good people being ruined because of things that just happen to them," but "something more deeply disturbing: it shows good people doing bad things." I am not suggesting that we need to be Greek heroes to be 'good' in the sense required to exemplify this sort of case. We need not be kings or military leaders, or be otherwise of elite status in any way, with illustrious lives filled with great successes and accomplishments. Nor, as I have tried to show, need we always do what we have undefeated reason to do. Tragedy may befall us when we fail to properly act responsively to reasons, just so long as those failures do not reflect badly on us as responders to reasons. We need only be 'good people' in a rather more quotidian sense. Something like this sense will be familiar from more colloquial discourse, as when we introduce a new colleague: "Meet the team. They're good people." When this is sincere, chances are the evaluation meant is a pretty down-to-earth one. We do not mean anything so lofty as that each of our colleagues is a living hero or a paragon of practical wisdom. We mean that they are just like the rest of us: ordinary, decent folk, who may sometimes make mistakes, but whose mistakes are not typically such as to reflect badly on them. They, like the rest of us, tend to go about their work without fault.

When we see Greek tragedy as concerned with laying bare the possible fates of ‘good people’ in this mundane sense, Nussbaum suggests (1994: 64), “we can far more easily see what a citizen would find terrifying here.” I am happy to admit that the chapters that follow are helpfully regarded as the work of a terrified citizen. I will be preoccupied with tragedy as it figures in our everyday affairs: *this* is the face of contingency I am concerned with throughout. We already encountered this face in subsection 1.3.3, perhaps most obviously where the father’s harming the family dog turns out to be non-negligent, since he was justified in believing that the dog was safe on the strength of what his app showed. We catch another glimpse with Williams’s (1979: 124) classic lorry driver case, “who, through no fault of his, runs over a child.” But to see the workings of tragedy in everyday life, which is what really interests me, we do not need to contrive examples in the philosophical imaginary. The figures I am most interested in, whose interests animate this dissertation from between the lines, should be familiar from the bewildering conditions of urban life in today’s political-economic order. I give due consideration to real-life cases throughout the dissertation, starting with the next chapter.²³

For now, what matters is that tragedy portrays people who are faultless in what they do, but who are intuitively apt to feel responsible regardless. Given what I said in section 1.3, the unique problem with tragic cases will, I hope, be clear. Whilst the Fault Principle is able to vindicate the phenomenology for the cases of resultant luck from subsection 1.3.4, it is not obviously able to do so for tragic cases, for the simple reason that the Fault Principle conditions the aptness of feeling responsible on being at fault, which the victims of tragedy are not. So, the unique problem comes to this: tragic cases allow the original dilemma to reassert itself, despite the sophisticated notion of ‘control’ encoded in our updated Control Principle (i.e. the Fault Principle). This is why I think tragedy is the face of contingency that confronts ethical thought with its deepest dilemma. Should we put our faith in the phenomenology and dispense with the Fault Principle altogether? Or should we hold fast to the ethical significance of faultlessness, of what is beyond our ‘control,’ and instead explain away the phenomenology as illusory? In the next chapter, we will see just how deep the dilemma goes, when we see the strange ways ethical thought contorts itself in attempting to respond to it. Needless to say, none of these shapes are especially attractive.

²³ In addition to Chapter Two, real-life cases are discussed at length in Chapters Six and Seven.

1.5. Concluding Remarks

In this chapter, I explored the idea that the contingency of the world can push life events beyond our control. There are phenomenological grounds for thinking that these include not just events that merely happen to us, but events we play an active role in, and for which we are responsible. Some have taken umbrage with the latter possibility, the possibility of moral luck, on the grounds that it is inconsistent with the highly plausible Control Principle. Moral luck cases thus appear to confront us with a dilemma: deny the phenomenology or deny the Control Principle. I argued that, for a range of cases, we can accommodate the phenomenology without denying the Control Principle, as long as we understand why it should be recast as the Fault Principle. I then showed that, for cases with a tragic structure, the dilemma nonetheless remains.

The project of the dissertation is to explain how ethical thought should respond to this most troubling face of contingency, to tragedy, in our lives. Chapter Two illustrates what is at stake in the dilemma about tragedy, laying out the main options as to how we can respond, as well as the shortcomings of each option. To anticipate, my preferred response requires that we account for a novel feeling of responsibility, one that is alike agent-regret in being apt in the absence of fault, but also alike blame in being able to do justice to the phenomenology of tragedy. Part Two goes on to provide such an account, which I carry forward into Part Three, where, as stated in the preamble, the discussion shifts into a political register.

CHAPTER TWO

Against Moralism about Obligation

2.1. Everyday Tragedy

In 2014, the International Federation for Human Rights (FIDH) registered a complaint to the European Committee of Social Rights against the Government of Ireland.¹ *FIDH v. Ireland* details a decline in standards of housing in several council estates in Ireland, since the 1980s, to the point of failure to meet the requirements of adequacy, suitability and habitability as set out in the European Social Charter and the Revised Charter. One of these estates was Dolphin House, under the authority of Dublin City Council (DCC). According to the complaint, the tenants of Dolphin House suffered “sewage invasions” of “grey and black wastewater,” later analysed to be “highly polluted,” which resulted in “dampness and fungal contamination” by colonies of *Aspergillus Fumigatus*, a “known human pathogen.”²

It is a deeply held conviction that no one should have to live in such conditions—least of all children. If our children have to tolerate such conditions, we may, as parents, feel deeply responsible for failing them. This seems to be the situation of one anonymous

¹ Complaint No. 110/2014 *International Federation for Human Rights (FIDH) v. Ireland*.

² *FIDH v. Ireland*: 21.

Dolphin House tenant, who poignantly remarks, “I must be a terrible parent to leave my children living in these conditions which I know are causing their ill health.”³ We can, I believe, hear behind this person’s words a feeling of responsibility for what they see as their failure to safely house their own children, something they take it they are obligated to do. Let us suppose, with the tenant, that they have indeed failed in this respect.

I think it goes without saying that the tenant’s failure is not their fault. Consider that the derelict state of Dolphin House was but a symptom of a wider malaise in social policy that would frustrate the regeneration plans of many local authorities. The story with DCC goes like this. *FIDH v. Ireland* details that “from 2001 onwards, Public Private Partnerships (PPPs) were promoted as the principal mechanism to create regeneration,” releasing funds by “land transfer arrangements between local authorities and private developers, alongside private investment.”⁴ As with many local authorities, this policy would leave DCC dangerously exposed to the whims of global financial markets. Thus, when, following the 2008 crisis, “the private residential and commercial aspects [of DCC regeneration plans] were no longer deemed economically viable by private finance and all of the PPP projects collapsed,” many council tenants were left living “in substandard housing conditions,” including those at Dolphin House.⁵ I hope we can agree that none of this is the tenants’ fault.⁶

This is a case of everyday *tragedy* as described in the previous chapter. The tenant’s feelings of responsibility are not just intelligible, but genuinely apt, *despite* the faultless nature of the failure to which they respond. This seems to be how the tenant sees things, and it is how we should see them too. Or so I want to suggest. The difficulty, as we saw in the last chapter, is that the phenomenology of tragedy is in tension with the intuitive Control Principle, even in its refined Fault Principle form. We want to do justice to the phenomenology expressed by the tenant, but equally, we want to retain commitment to

³ *FIDH v. Ireland*: 35. The interview was conducted for a report by the Rialto Rights in Action Group, a local grassroots organisation.

⁴ *FIDH v. Ireland*: 38.

⁵ *FIDH v. Ireland*: 39.

⁶ This impression is only strengthened by the recognition that PPPs, despite the rhetoric of harnessing private sector dynamism to achieve public value for money, are basically just there to aid the narrow electoral interests of incumbent governments, by allowing public services and infrastructure to be financed off the books. See Rory Hearne (2014: 160-7) for how the story in DCC’s specific case tends to corroborate this point vis-à-vis the Irish government at the time.

the intuitive Fault Principle. In this chapter, we explore a debate about the coherence of tragic cases, showing how the depth of this dilemma makes itself felt in just how hard it is to come to a settled position. I begin by presenting the debate in question.

2.2. The Debate over ‘Moralism’

An easy way to question the coherence of tragic cases is to consider how, in the Dolphin House case, the idea that the tenant could be failing as a parent might have struck us as somewhat unfair, given the circumstances. After all, as we said, the tenant was without fault given the circumstances. And if, as we know, that means they either did what they had to do, or at least what could be reasonably expected of them, then how could it make sense to say they failed to live up to their parental obligations? Since this skeptical view implies that those who conduct themselves faultlessly (a saint; a *phronimos*) cannot be in breach of obligation, I call it ‘moralism about obligation,’ or ‘moralism’ for short.

Whatever we think about moralism, it has to be taken seriously because there is a powerful argument in its favour, which I call the MORALISTIC ARGUMENT. This argument aims to establish what the moralist is after: a conceptual link between fault and breach of obligation. It is the strongest and most succinct argument for moralism of which I am aware.

MORALISTIC ARGUMENT

- (1) Situations of faultless failure are possible, where people both
 - (1a) Stand in breach of an obligation,
 - (1b) But are not at fault in so doing.

{assume for reductio}
- (2) Anyone who breaches an obligation is thereby blameworthy for so doing.
- (3) No one is blameworthy for anything without being at fault in so doing.
- (4) A person, who faultlessly fails, stands in breach of an obligation.

{from 1a}
- (5) The same person does not stand in breach of that very obligation.

{from 1b, 2, 3}

(6) Since (4) and (5) are contradictory, (1) must be false.

{from 2, 3, 4, 5}⁷

The argument takes the form of a valid *reductio ad absurdum*, and its power lies in the plausibility of its main conceptual assumptions: premisses (2) and (3). Premiss (2) says that *breach of obligation entails blameworthiness*, or *BEB*, for short; premiss (3) says that *blameworthiness entails fault*, or, as we will say, *BEF*. When formalised this way, we can see clearly how moralism takes a stance on the moral luck dilemma as it arises for tragic cases. For *BEF*, as we can see, is the Fault Principle. So, faced with the question of whether to deny the phenomenology in tragic cases or to deny the Fault Principle, the moralist cannot but take the first option. Of course, if the dilemma is worth its salt, this will not be a costless exercise—and, indeed, the dilemma is worth its salt. The moralist, as we shall see, is forced to respond to the Dolphin House tenant in demonstrably unsavoury ways. There are kinder and meaner ways this could go, but we should want to avoid them both. Let me explain.

The kinder moralist holds fast to the tenant's faultlessness. But since, *qua* moralist, they deny the coherence of faultless failure, they are forced to deny that there is in fact any failure on the tenant's part that they could be apt to feel responsible for at all. So, paradoxically, this approach is 'kind' to the tenant by pathologising their legitimate feelings. The ethical ramifications of doing so become especially urgent in the context of mental health services. According to poverty research by Felicity Thomas, Katrina Wyatt and Lorraine Hansford, general practitioners (GPs) in the UK often medicalise the distress their patients feel when they are not doing very well, including by their loved ones. The study shows that GPs treat "mental distress as caused by [...] pathological issues related to brain 'dysfunction,'" which "in turn leads to responses that promote what are often ineffective and potentially harmful medical treatments in the form of antidepressant drugs" (2020: 1133). Sometimes this stems from ignorance about "the

⁷ Something like this argument is anticipated by Ruth Barcan Marcus (1980: 126-7), who says: "If an agent ought to *x*, then he is guilty if he fails to do it. But if, however strong his character and however good his will and intentions, meeting other equally weighted or overriding obligations precludes his doing *x*, then we cannot assign guilt, and, if we cannot, then it is incoherent to suppose that there is an obligation. Attendant feelings of the agent are seen as misplaced."

inequalities that cause this distress in the first place,” and sometimes simply from “the pressures faced to help a patient within the short timeslot of a consultation” (ibid.).⁸ Either way, they make vivid the problem with the ‘kinder’ moralism. Not only does it misdiagnose, and so maltreat, the patients’ feelings. It also occludes the normative reality to which those feelings respond, and with it the “social and structural determinants” (ibid.) of how the patients fare in that reality.

Yet more paradoxically, the moralist who takes the tenant’s feelings seriously turns out to be the meaner moralist. Since, *qua* moralist, the aptness of the feelings cannot be pinned on a faultless failure, this approach is forced to conclude that the failure must in fact have been the tenant’s fault.⁹ We are reminded of the former UK Prime Minister Margaret Thatcher, who in 1978 said that if people experience poverty, it is because they “don’t know how to budget, don’t know how to spend their earnings,” an ignorance expressive of a “fundamental character-personality defect.”¹⁰ This logic is most visible in access to institutional support, both statutory and third-sectoral. For example, Thomas, Wyatt and Hansford (2020: 1128) found that amongst street-level bureaucrats in both types of agencies, service-users “were seen as witting players in their own plight; [...] that they were ‘not bothering to help themselves’; that their situation of precarity was directly linked to their ‘chaotic’ and ‘irresponsible’ lifestyle.” Such attitudes were in turn manifest in punitive and often downright bizarre assessments, as when for one service-user “a problem with mould in her flat had been put down to her cooking the ‘wrong’ type of food, and placing her furniture against the walls in an attempt to maximise space” (ibid.). To try to find fault in the Dolphin House tenant in this way would be, I believe, woefully misguided.

Thus, moralism seems to force one of two troubling responses: *either* we pathologise the tenant’s feelings *or* we strain to find fault in their conduct. What I find troubling is

⁸ Perhaps surprisingly, it is interviews with GPs who medicalise on the latter basis that seem the more troubling. For instance, one GP diagnoses their patient with depression and anxiety, whilst admitting that “I know in my heart of hearts that it’s not a medical problem.” Another says that by prescribing antidepressants, “you feel that at least you tried to give something when [...] you can’t give them a roof over their heads, you can’t change the fact that they don’t have any support or family around” (ibid.).

⁹ See Chapter Seven for discussion of a moral-metaphysical rationale for this view.

¹⁰ Interview of December 22nd 1978 by Richard Dowden for the *Catholic Herald*, available at <https://www.margaretthatcher.org/document/103793>. (Accessed April 26th 2024.)

not just how these responses treat the tenant. It is also that they betray a fundamental misapprehension of what is going on, one I suspect, with Thomas, Wyatt and Hansford (2020: 1133), works to “inadvertently depoliticise suffering and in turn, reinforce and naturalise the very structures that uphold and reproduce oppression.” So, there are not only ethical reasons to oppose moralism but possibly political-epistemological ones too. Seen this way, the question is not *whether* to oppose moralism, but *how*.

If the MORALISTIC ARGUMENT is anything to go by, there are basically two forms our opposition to moralism could take: we could deny *BEB* or *BEF*. Start with the denial of *BEF*. Since, as we saw, *BEF* is the Fault Principle, we can understand the anti-moralist who denies *BEF* as opting to resolve the dilemma in favour of the phenomenology. Here is how that works with the Dolphin House tenant. Firstly, via *BEB*, they are interpreted as aptly blaming themselves for their parental failure, and then, by denying *BEF*, that blameworthiness is allowed to stand despite the absence of fault. In this way, this form of anti-moralism does justice to the tenant’s phenomenology, instead of pathologising their response to the situation, or, worse yet, redescribing the situation as somehow their own fault.

Responses to tragedy in the literature have often expressed an anti-moralism of this form. An early example is Ruth Barcan Marcus (1980: 130), who takes the moralist’s response to tragedy “to claim that it would be mistaken to feel guilt or remorse about having failed to act according to that obligation [...] since [the] obligation was vitiated and [our] feelings are inappropriate,” a claim which, she continues, “is false to the facts.” Having thus endorsed *BEB*, she goes on to claim that there are at least some obligations that may be breached “however holy our wills or rational our strategies” (1980: 135), that is, however faultless we may be.¹¹ Another example is Martha Nussbaum, who asks if Agamemnon’s sacrifice is a case of “actual blameworthy wrongdoing” (2001: 27), only to find that his downfall speaks “to our intuitive demand that even the constrained killer should come to regard himself as a killer and should suffer, in his own person, for his deed” (2001: 41), a demand reflecting “a deep ethical response that would be intelligible in the absence of the divine” (ibid.). More recently, Amy Sepinwall (2017: 524-5) nods approvingly to aspects of “the Ancients’ responsibility practice,” seeking to recover “the

¹¹ For Marcus (1980: 134-5), the prime example is the higher-order obligation to act in ways that avoid conflicts of obligation: situations that, if we act with justification, would be a subset of tragic cases in our terms.

rationale for this more expansive responsibility practice,” within which “one can be *blameworthy* even if one is not at fault.” In a similar proposal, Elinor Mason (2019: 248-9) considers someone who inadvertently loses her friend’s sentimental necklace, finding that “it is plausible that she should feel something in the region of remorse,” and that she “should feel really bad about what she did, *even though* there was no bad will.”

Despite the popularity of this form of anti-moralism, I believe it leaves much to be desired, at least given the reasons I gave above to oppose moralism. By insisting that the Dolphin House tenant is to blame for their failure despite their lack of fault, it seems the present view treats them no less unpleasantly than does the moralism it seeks to oppose. For this reason, I think the denial of *BEF* is sensibly labelled ‘hard’ anti-moralism. We should want to avoid moralism, but seen this way, it is not clear that hard anti-moralism is any better. So, if it turns out that this is the only way to be an anti-moralist, we may, however begrudgingly, begin to feel tempted by moralism. But it is hardly a stable resting place. Neither moralism nor hard anti-moralism seem easy to swallow, yet we are forced to choose. The confusion may well have an air of familiarity by now: it merely expresses the depth of the moral luck dilemma as it arises for tragic cases, which we encountered in the previous chapter.

Fortunately, as we saw, the MORALISTIC ARGUMENT makes available another route to anti-moralism which I call ‘soft’ anti-moralism. Soft anti-moralism proceeds by denying premiss (2) instead of (3), *BEB* and not *BEF*. In the context of our debate, it appears to be something of a silver bullet. On the one hand, it can accept that the Dolphin House tenant is faultless without the hard anti-moralist suggestion that they are nevertheless blameworthy. But on the other hand, it does not moralistically pathologise the tenant’s feelings of responsibility for their failure. We may wonder how soft anti-moralism achieves this if it denies *BEB*. The trick is to reinterpret the tenant’s phenomenology: though what they feel is perfectly apt, it is not in fact any form of blame but rather the quite distinct feeling of *agent-regret*. Agent-regret is what Bernard Williams says his lorry driver is apt to feel, having faultlessly run over a child, a sentiment he famously says it would be “a kind of insanity” (1976: 125) for them not to experience. The appeal to agent-regret, as we will see, is not without its challenges. But for now, it is enough to register that soft anti-moralism allows us to oppose moralism without denying *BEF*.

The debate can be summarised as follows. We should all want to be anti-moralists, but only if its ‘soft’ form is available. So, the question is: does the MORALISTIC ARGUMENT

go through, and if not, can we accept soft rather than hard anti-moralism? In this chapter, I aim to establish that the MORALISTIC ARGUMENT fails at *BEB* and not *BEF*, making soft anti-moralism not only the most desirable but the only defensible position on the debate. I defend *BEF* in section 2.3 and argue against *BEB* in section 2.4. In section 2.5 I go on to offer a philosophical diagnosis as to why so many anti-moralists have been led to deny the wrong premiss, revealing a key moral-psychological desideratum that any viable soft anti-moralism will need to meet. Section 2.6 briefly concludes and looks ahead to Part Two.

2.3. Blameworthiness Entails Fault

Denying that blameworthiness entails fault, *BEF*, is the essence of hard anti-moralism. In the literature, *BEF* is usually denied on the strength of the phenomenology in tragic cases, rather than that of any argument against *BEF* itself. Of the hard anti-moralists mentioned, it is probably Sepinwall (2017) who has gone the furthest in arguing the case against *BEF*. To the extent that the others offer arguments independent of the phenomenology, they are similar to Sepinwall's. I therefore focus on her arguments, with a view, ultimately, to rejecting them.¹²

In moral and legal philosophy, *BEF* is usually called the 'Fault Principle.' The Fault Principle is intuitively very plausible and seems thoroughly entrenched in our blaming practices, in both their interpersonal and institutionalised forms.¹³ Perhaps the strongest indication of this is its appearance in the American Law Institute's (1985) *Model Penal Code*, which at §1.02(1)c states that one of its guiding purposes is "to safeguard conduct that is without fault from condemnation as criminal." The Model Penal Code has been enormously influential, and since its first publication in 1962 it has played a significant role in the recodification of criminal law in a majority of U.S. jurisdictions.¹⁴ We may therefore have our doubts about the wisdom of denying the Fault Principle. The point is

¹² This is so despite my sympathy, not only for her anti-moralism (though not its 'hardness'), but also for some of her political motivations (2017: 527).

¹³ See Chapter One for my own defence of the Fault Principle.

¹⁴ For this and other claims about the influence of the Model Penal Code on black-letter law, see Markus Dubber (2015: 5-7).

not that denying it would have highly revisionary real-world implications: though that could show our doing so to be ill-advised, it could never bear on the truth or falsity of the Fault Principle. The point is rather denying it would go against the verdict of many relevantly well-credentialed people, which really might be good evidence for its truth.

Depending on our views, however, the usual relationship between expert testimony and evidentiary status may not be thought to apply in the present case. For instance, if we identify the relevant demographic of experts (i.e. one comprised of academics, jurors, lawmakers, and so on) as some sort of an Establishment with its own partial interests, then the sociological fact that the Fault Principle enjoys expert endorsement may, in the presence of certain political predilections, actually count against it. Given that the Fault Principle rarely receives anything approaching explicit defence amongst moral and legal philosophers,¹⁵ a healthy degree of skepticism on these sorts of grounds is not entirely unreasonable.

This is roughly where Sepinwall stands on the Fault Principle. She is one of the few to have noticed the strange paucity of argumentation around the Fault Principle,¹⁶ and one of even fewer to have developed arguments against it. Her basic claim is that though fault may be sufficient for blameworthiness, it is not in fact necessary. For someone to be blameworthy, she suggests, all that is needed is a *good moral reason* to blame them, which may but need not be the fact that they are at fault. In her view, then, non-fault moral reasons can just as well make someone blameworthy as fault-based moral reasons can (2017: 529).

This should sound familiar from consequentialist accounts of blame, which all say,

¹⁵ Though the Fault Principle often receives explicit thematisation in criminal law theory, it usually gets assumed on the strength of scholarly consensus, as when Stephen J. Morse (2000: 879) says that “virtually all criminal law theorists agree that moral fault is at least a necessary condition of blame and punishment.” In moral theory, it may be lucky to get recognised as an assumption at all. For instance, Susan Wolf (2001: 16) ends up partially agreeing with what she calls the ‘rationalist’ view that “blameworthiness is solely a function of faultiness.” But Wolf interrogates rationalism mostly by pitting it against its ‘irrationalist’ rival, unperturbed by how rationalism gets to make its central claim in the first place. As a result, the Fault Principle never comes into full view as an assumption. When moral theorists do recognise the Fault Principle, it tends to get assumed on the strength of intuition, as Miranda Fricker (2016: 168-70) does in her paradigm-based vindication of our blaming practices.

¹⁶ This is a merit she shares with Gardner, who, in contrast to Sepinwall, offers an intriguing argument in its favour (2005: 124-8).

typically hedged in various ways, that it is necessary for the aptness of (a system of) blame that it maximises good consequences.¹⁷ It is instructive to consider why such accounts are thought to fail. The standard explanation is that they require us to blame the faultless if this happens to maximise the good, which David Boonin (2008: 41) helpfully calls the “punishing the innocent objection.” When asked why this seems an obvious injustice, the natural answer is to plead the Fault Principle—though that, of course, risks begging the question in the present context. But it is not question-begging to say that maximising the good strikes us the *wrong kind of reason* to treat someone punitively, whether through blame or more institutionalised means.

To see the distinction between the right kind of reasons (RKR) and the wrong kind of reasons (WKR), consider feelings like fear or joy.¹⁸ In one sense of ‘should,’ we should only fear what is truly fearsome, or rejoice what is truly joyous. These strike us as RKR for fear or joy, making them *apt* or *fitting* feelings, which we ‘should’ in that sense feel. But in another sense we ‘should’ sometimes fear what is quite innocuous, or rejoice what is quite mundane. Maybe it would heighten a younger sibling’s enjoyment of the haunted house if they thought we were afraid, and the easiest way to achieve this would be if we actually feared the underpaid student in zombie attire. Or maybe it would keep our clients sweet if they thought the corporate soirée was a raging success, and the only way to believably convey this is if we somehow actually enjoyed the occasion. In these cases, there really are reasons to feel fear or joy, but they clearly do not make these feelings apt or fitting—they strike us as the WKR for that. They speak to the broader *justifiability* of feeling fear or joy, fittingly or otherwise. Being fitting and being justified are therefore independent modes of propriety: it may sometimes be justified to feel unfitting feelings, or unjustified to feel fitting feelings.

¹⁷ The basic rationale, in Jeremy Bentham’s famous words, is that “Upon the principle of utility, if [punishment] ought at all to be admitted, it ought only to be admitted in as far as it promises to exclude some greater evil” (2007: XIII, §1.II).

¹⁸ The classic statement of this distinction is due to Justin D’Arms and Daniel Jacobson (2000). Some, such as Jonathan Way (2012), are skeptical of the distinction between RKR and WKR in general because they think all putative WKR for *S* to *x* are in fact RKR for *S* to want or to make it the case that they *x*. Others, such as Barry Maguire (2017), are skeptical of the idea that the distinctive normativity of facts that make our reactions fitting is the normativity of reasons, so that calling them ‘RKR’ is misleading. Though I am not averse to these views, they remain a matter of debate, so I proceed on the assumption that the distinction is sound.

When we blame someone for doing something ‘worthy’ of it, is the propriety thus conveyed that of fittingness or justification? Or, to say the same, is it RKR or WKR in light of which a blameworthy person should be blamed? There is a simple test for this. Imagine that someone is blamed for something, but that circumstances obtain such that they are not blameworthy for what they are being blamed for. Now, holding fixed these circumstances, imagine that an evil demon contrives to destroy the world unless we also blame this person. That it is the only way to avert the destruction of the world justifies our blaming this poor person if anything does, albeit for a paradigm WKR. But clearly, however, that does not make them blameworthy. This suggests the following criterion of adequacy for an account of blame: what is said to make people ‘worthy’ of it must be an RKR.

The question for Sepinwall is whether the non-fault moral reasons alleged to make people blameworthy are RKRs for blame. As we will see, the relevant non-fault moral reasons either are or are explained by the fact of some relationship that obtains between the putatively blameworthy person and someone who is injured by their conduct, or as she says, “relationship-based reasons” (2017: 525). So, are these relationship-based reasons plausibly RKRs for blame? Sepinwall has three arguments to the effect that relationship-based reasons can influence the fittingness of blame independently of fault. I argue that none is satisfactory, for on closer inspection the relevant reasons turn out to be either *merely apparent* RKRs or *bona fide* RKRs but for the *wrong thing*.

The first argument runs as follows. Our standing in certain types of relationships has a special demandingness. When we injure those to whom we stand so related, this special demandingness can make it reasonable—at least from our own perspective—to believe that we were at fault, even if in fact we were not (2017: 546-7). In such cases, Sepinwall claims, self-directed blame (e.g. feelings of remorse or guilt) may be a genuinely fitting response to our belief that we were at fault. This doxastic fact is the candidate relationship-based reason. It is said to be an RKR for self-blame, one that may obtain even where we are not at fault.

The problem, I argue, is that the fact we believe we were at fault is a *merely apparent* RKR to blame ourselves, but what is needed is a *bona fide* RKR. Our fitting feelings are in this sense veridical. To see this, consider an epistemological analogy. For a belief that *p* to count as knowledge it needs to have been formed on the basis of evidence—an RKR for belief—rather than a WKR, such as that an evil demon compels us to believe that *p*.

Merely apparent evidence, though not straightforwardly a WKR, is not enough to make a belief count as knowledge. Much of what the average Soviet citizen believed about their regime from reading *Pravda* was not knowledge. To count as such, the reportage at *Pravda* would have to have been transparent to what was really going on in the Soviet Union, which, of course, it often was not. Analogously, if self-blame befits a belief that we were at fault, it is only if and because the belief is transparent to the fact that we were at fault: the RKR has to be more than just apparent. Otherwise, our self-blame is at best intelligible, but falls short of fittingness. This is not to deny that the fact that we believe we were at fault may sometimes be a bona fide reason of *some* kind to self-blame. Maybe an evil demon threatens to destroy the world unless we blame ourselves in response to such a belief, no matter if it is true. But of course, if the relevant doxastic fact is a reason of *this* kind, it is a WKR: though it may justify self-blame, it never makes it fitting.

Sepinwall's second argument starts with the following observation. When someone injures another, their degree of blameworthiness depends on the type of relationship they stand in with their victim, even when we hold fixed: (a) the injury, and (b) the degree of fault. To see this, imagine two equally negligent drivers hit and injure a child. If one is the child's parent, and the other merely a stranger, we might expect the former to blame themselves more than the latter. Sepinwall infers from this that there must be fault-independent relationship-based considerations that determine how much blame we are apt to feel (2017: 549-50). Since these non-fault considerations bear in some way on the aptness of blame, it would seem that they RKRs for blame.

I agree that non-fault considerations can indeed determine how much blame is apt. There are many such considerations, including the one Sepinwall highlights: the type of relationship that obtains between injurer and victim.¹⁹ Quite clearly, the significance of injuring our own children differs from injuring a stranger's. This is likely explained by the differing value of our relationships to loved ones versus strangers, and the differing patterns of normative expectations that partly constitute them. These relationship-based considerations may well be genuine RKRs for blaming more or less, in a way that, say, reasons of political expediency are probably not. But that is not really to the point. What Sepinwall needs is to provide bona fide non-fault RKRs that tell us not *how much* blame

¹⁹ Others might include considerations of proportionality, the relationship between judger and injurer (including their relative 'standing'), and whether and to what extent the injurer is independently remorseful.

is apt, but that blame is apt *in the first place*. The relationship-based reasons she highlights do not plausibly play this role. So, whilst I agree that how much we blame a negligent injurer may fittingly vary with their relationship to the victim, whether it is fitting to blame them *per se*—whether they are blameworthy—is surely not something that varies with the relationship. It is blameworthy to injure negligently, no matter the relationship between injurer and victim.

The core thought behind Sepinwall's third and final argument is as follows (2017: 551-2). When we injure those to whom we stand in certain relationships, their special demandingness can turn the mere fact of our causal involvement into a *prima facie* reason to blame ourselves. Since we may be causally involved in injuring the victim without being at fault, the fact of our causal involvement is said to be, at least sometimes, a non-fault RKR for self-blame.

An initial difficulty is that the feeling usually taken to befit such a fact is agent-regret, rather than any form of self-blame (e.g. feelings of remorse or guilt). Sepinwall's first move is to suggest that agent-regret *just is* a form of self-blame, saying that "there is no meaningful difference between agent-regret and guilt" (2017: 551). We might worry that this only secures a bona fide non-fault RKR for blame by denying a widely accepted and phenomenologically robust distinction. But even putting this to one side, we might still wonder whether a *prima facie* reason for self-blame suffices to secure blameworthiness. After all, it may be that the *prima facie* reason is defeated by the finding that, though we were indeed causally implicated, we were not at fault. Recognising this, Sepinwall further suggests that the special demandingness of the relationship not only makes mere causal involvement a *prima facie* reason for self-blame, but also a reason not to *enquire into* whether we were at fault. Here, she leans on the idea that any such preoccupation with exculpation would disrespect our relationship with the victim. This is a nice idea, but it cannot help Sepinwall, for if the candidate reason is an RKR at all then it bears on enquiring into fault rather than on self-blame. It may seem that the idea can be salvaged by appealing to the fact that if we heed these reasons not to enquire into fault, we remain ignorant as to what exactly happened. So if, as per the first argument, the relationship makes our epistemic situation one where we reasonably believe we were at fault, then perhaps this doxastic fact could serve as the candidate RKR? But if this is the proposal, it runs headlong into the objection from before, viz. that such a reason can only be an apparent rather than a bona fide RKR. In sum, if Sepinwall's third argument yields novel

bona fide RKR at all, they are RKRs for the wrong thing.

I conclude that, for all their ingenuity, Sepinwall's arguments do not succeed. Where does this leave us? Of the hard anti-moralists, it is Sepinwall who has done most to argue against the Fault Principle, and thus to deny *BEF*. If her arguments fail, as I claim they do, the denial of *BEF* becomes unsustainable, and with it, hard anti-moralism itself. On its own, of course, this is grist to the moralist's mill. In the next section, I make the soft anti-moralist case that the MORALISTIC ARGUMENT fails at *BEB*, even if not at *BEF*.

2.4. Breach Does *Not* Entail Blameworthiness

My case against the claim that breach of obligation entails blameworthiness, or *BEB*, will proceed by counterexample. Before we begin, we should get clear on what actually *BEB* says and where it could have come from. Perhaps the most familiar expression of the idea in the Western tradition is owed to John Stuart Mill (2015: 161-2), who puts it this way:

We do not call anything wrong, unless we mean to imply that a person ought to be punished in some way or other for doing it; if not by law, by the opinion of his fellow-creatures; if not by opinion, by the reproaches of his own conscience. [...] It is part of the notion of Duty in every one of its forms, that a person may rightfully be compelled to fulfil it.²⁰

In this passage, Mill speaks of obligation, or as he says 'duty,' in some sense that warrants honouring it with a capital 'D.' Presumably, by Duty (capital 'D') he intends the

²⁰ By gesturing in Mill's direction, I do not mean to suggest that *BEB* is a characteristically consequentialist commitment. For example, consider Kurt Baier (1966: 223), who writes, "What [...] gives directives morally binding force [...] is that they concern themselves with issues and problems whose solution is *not solely the agent's business*," where this means "it is justifiable for a society to take suitable measures to ensure that its members follow them," including "pressures which are in themselves obnoxious." More recently Stephen Darwall (2006: 93) writes, "We hesitate to impute wrongdoing unless we take ourselves to be in the range of the culpable." If anything, *BEB* is characteristic not of consequentialism but rather the metaethical view David Owens calls the 'sanction theory of obligation' (2008: 404-8).

output of his preferred utilitarian calculus, as distinct from whatever ‘duties’ (lower case) we take ourselves to have in ordinary non-utilitarian consciousness. The capital ‘D’ indicates that Mill imagines there to be—to speak metaphorically for a moment—duties or obligations at two different levels, to which correspond two different senses of ‘duty’- or ‘obligation’-talk. As it happens, Mill, as we can see, gave expression to *BEB* in terms of ‘obligation’ in the higher sense. But nothing in principle prevents it from being couched in the lower sense. Regardless of how Mill intended *BEB*, the key takeaway is that *BEB* admits of different readings depending on the sense of ‘obligation’ involved.²¹ There are several ways we could mark the distinction nowadays, but perhaps the cleanest is Jonathan Dancy’s. To have an ‘obligation’ in the lower sense is a fact at what Dancy would call the *contributory* level, defeasible as a reason for action, whereas to have an ‘obligation’ in the higher sense is undefeated as a reason for action, making it a fact at the *overall* level (2004: 15-6). This will become important as the argument progresses.

The first counterexample to *BEB* I want to consider is where breach of obligation is adequately excused. Most accounts of excuses accept the following two criteria: (a) they presuppose a failure to satisfy some reason or other (which may but need not be limited to obligations), and (b) they are exculpatory when adequate, showing that the failure is not blameworthy.²² Whatever we may think of how different accounts go on to explain these criteria, or how excuses function in human life, the intuitiveness of the criteria is, I believe, both bedrock and pervasive.²³

For instance, someone who misses a bidding deadline for a rare item they wanted on eBay might beat themselves up over it (‘How could you miss the deadline? You really wanted that thing!’), but if they missed it because the Wi-Fi went down, it would be

²¹ Richard Brandt, in an early paper, applies sustained scrutiny to the claim we are calling *BEB*, and does well not only to recognise its ambiguity, but to keep the two senses of ‘obligation’ well apart during the course of his critique (1958: 34-9).

²² Note that John Austin may well be an exception here. According to Austin, when someone has an excuse for doing something they “admit that it was bad but don’t accept full, or even any, responsibility,” so that “it is not quite fair or correct to say *baldly* ‘[they] did [it]’” (1957: 2). One way of reading this is that excuses, for Austin, require (b) but not (a). Sometimes R. Jay Wallace seems to follow this reading of Austin, when he says things like “excuses serve to show that an agent has not really done anything wrong” (1994: 127). For a powerful argument against this type of account, see Michael Zimmerman (2004).

²³ What follows is consistent with the more detailed treatment of excuses in Chapter One.

familiar and intelligible that they cut themselves some slack, and we would understand this using the concept of an excuse. Even such trivial cases exhibit the need for the two criteria. Had they relaxed their self-disapproval in response to something they did that was in no way regrettable (i.e. were (a) absent but (b) present), then this behaviour would not be intelligible under the concept of an excuse. Similarly, had they responded to missing the eBay deadline in any way such that they still beat themselves up over it (i.e. were (a) present but (b) absent), then they could not be understood as excusing themselves either. Recognisably the same concept (with the same criteria) applies to less trivial examples from everyday life, where the relevant failures strike us as breaches of obligation: for example, if we replace missing the eBay deadline with the breaking of an important promise. And again, recognisably the same concept is institutionalised in the criminal law of both common and civil law jurisdictions, through the practice of excusatory defences. Since only pleas that presuppose legal responsibility for an offence count as defences, excusatory defences satisfy criterion (a).²⁴ And since they would not count as defences if they did not get defendants off the hook of criminal liability when successful, they also satisfy criterion (b).

These examples show that we not only frequently take ourselves to encounter cases of excused breach of obligation in ordinary life, but that we are also not wrong to think of such cases in these terms. This makes it hard to deny the possibility of excused breach of obligation. Given that we are blameless for breaches of obligations for which we have an adequate excuse, it is just as hard to deny the possibility of cases of breach without blameworthiness. For now, we treat these counterexamples to *BEB* as dispositive, though we shall return to them later on in this section.

The proponent of *BEB* might respond by seeking refuge in the distinction between the contributory and the overall. To see how, consider that excuses operate at the overall level. In general, when someone does something, it is only if there is undefeated reason for them not to have done it that they stand in need of an excuse. The same applies if what they did was in breach of obligation: it is only if the breached obligation happens to be undefeated—that is, if it lies at the overall level—that the question of excuses can sensibly arise. If the breached obligation lies only at the contributory level, so that it is

²⁴ Admittedly, some lawyers and legal theorists do call certain pleas ‘defences’ that do not satisfy criterion (a). See Antony Duff for reasons to think that that use of the term is misleading, and for a brief history and systematic justification of our preferred usage (2007: 263-4).

defeasible by other considerations, then its breach is not automatically the sort of thing to which excuses can apply. When an obligation is defeated by competing considerations, so that breach is justified overall, no excuse is necessary: we do not need an excuse for what we are justified in doing. The proponent of *BEB* may thus attempt to evade the counterexample by saying that ‘obligation,’ as it figures in the premiss, is meant in the contributory rather than the overall sense.

The problem with this manoeuvre is it faces counterexamples of its own, this time from *justified* breach of contributory obligation. We already leaned on the intuitive possibility of such cases in the last paragraph, and concrete examples are not too far to seek. Consider promises that are broken because something comparatively important arises, as when a firefighter on call breaks a promise to attend her boy’s birthday party due to an emergency response request citing immediate threat to life. Or consider the institutionalised practice of justificatory defence in criminal law procedure, the logic of which begins, as with all legal defences, by conceding the commission of an offence, and proceeds to argue that this was in fact justified under the circumstances.²⁵

Once we accept that such cases are possible, it is a short step to seeing that they are counterexamples to this reading of *BEB*. Sometimes, as we said, we have an undefeated reason to breach a contributory obligation. When we do so *for* that undefeated reason, we are justified in so doing, and that seems to get us off the blame hook.²⁶ Pressed to say why, the answer, in short, is that blame criticises how we fare as responders to reasons: it criticises that and how we failed to satisfy the balance of reasons confronting us.²⁷ So,

²⁵ For this view of justificatory defences in law, see, again, Duff (2007: 218n87).

²⁶ On the need not only for an undefeated reason, but to have acted for that reason, see John Gardner (2007a: 94n7), who refers us to the English legal precedent set in *R v Dadson* [1850] 2 Den. 35; 169 E.R. 407. In *Dadson*, a police officer was sentenced for shooting and wounding an escaping felon without knowing that they were an escaping felon at the time. The defence argued that the act was justifiable by the fact that the felon was escaping. The prosecutor agreed, but still rejected the defence because the act, though justifiable, was not done *for* the reason that could have justified it. This point is sometimes missed. For instance, Stephen Darwall (2006: 98) challenges us to “Try formulating an expression with which you might address a moral demand to someone. I doubt that you can find one that does not carry the implication that she has conclusive reason to do what you are demanding or reason not to have done what you are blaming her for.” The decision in *Dadson* seems to be just that: the officer *had* conclusive reason to shoot the felon, but is blamed for not doing so *for* that reason.

²⁷ This intuitive answer is elaborated in more detail in Chapter One.

if we acted in accordance with and for an undefeated reason, there is nothing to critique in this regard: we are not to blame. This makes justified breach of contributory obligation a straightforward counterexample to *BEB* on this reading.

At this stage, it seems fair to say that *BEB* is in serious jeopardy. Whichever way we read ‘obligation’ as it figures in the premiss, it seems to face decisive counterexamples. On an overall reading it faces cases of excused breach, and on a contributory reading it faces cases of justified breach. Unless there is some other reading that avoids these counterexamples, it looks like the MORALISTIC ARGUMENT fails at *BEB*. The question is: does the proponent of *BEB* have any other options left at this stage?

A final option returns, as promised, to a version of the overall reading. The obvious difficulty, of course, is how to with cases of excused breach. The basic proposal is to read a *ceteris paribus* clause into *BEB*, such that breach is blameworthy *other things equal*, where other things are *not* equal when (for example) the breach is adequately excused.²⁸ Thus modified, *BEB* is no longer threatened by cases of excused breach, as it no longer claims unqualified entailment from breach to blameworthiness. After all, it no longer denies that where there is an excuse the inference cannot be drawn.²⁹

The question for this proposal is how exactly the *ceteris paribus* clause is supposed to make the inference from breach to blameworthiness defeasible. There are two ways of understanding the defeasibility at issue. The first is to understand it consistently with what logicians would call the ‘monotonicity’ of entailment. On this approach, the *ceteris paribus* clause makes the inference defeasible because it expresses a suppressed conjunct in *BEB*’s antecedent, which if false prevents the inference from being drawn in the usual way. In other words, the suppressed conjunct is only true if there is no excuse.³⁰ To spell it out, with the *ceteris paribus* clause included, *BEB* now reads: *if* someone is in breach of obligation *and* unexcused, *then* they are blameworthy. Hence, for someone who *is* in breach but *has* an excuse, then the modified *BEB* straightforwardly does not license the

²⁸ I assume an *exclusive* conception of the relevant *ceteris paribus* clause, where other things are ‘equal’ when they are ‘right.’ I disregard the *comparative* alternative, since the counterfactual reasoning it would require involves us in claims about what should count as the nearest possible world, which it is unclear what could settle. See Gerhard Schurz (2002) for this distinction.

²⁹ For an example of the *ceteris paribus* reading of *BEB*, see David Owens (2016: 5).

³⁰ It makes no difference to the current objection whether or not what it is to lack an excuse is finitely specifiable, or as John Earman, John Roberts and Sheldon Smith would put it, whether or not the *ceteris paribus* clause is “a function of laziness” (2002: 283).

conclusion that they are blameworthy.

Though this modification to *BEB* looks promising, its fatal flaw is that it invalidates the MORALISTIC ARGUMENT. Recall that this argument is a *reductio* of the possibility of tragedy, of situations of faultless breach, with the contradiction being that those who find themselves in such situations both do and do not stand in breach of obligation. We draw the contradiction by first inferring that those who are faultless are blameless (by *BEF*), and then that they are not in breach (by *BEB*), thus contradicting the supposition that those in tragic situations are faultlessly in breach. On the modified *BEB*, however, though we can still infer that the faultless are blameless, it can now only be inferred that the blameless are not *both* in breach *and* unexcused. And this no longer contradicts the supposition that those who faultlessly fail are in breach: just consider that someone not both in breach and unexcused may be someone *in breach* but who is excused. Thus, this modification to *BEB* invalidates the MORALISTIC ARGUMENT. Of course, if the argument is invalid, then it becomes dialectically quite pointless to quarrel over the soundness or otherwise of its premisses. The various positions on the debate over moralism turn out to be ill-defined.

The second way to understand the defeasibility introduced by the *ceteris paribus* clause does away with the monotonicity of entailment. On this approach, the inference from breach to blameworthiness is defeasible by an excuse not because the antecedent harbours a suppressed conjunct, but because the inference itself may be retracted depending on the informational context. To illustrate the idea, imagine a list comprised of two premisses: that breach entails blameworthiness (i.e. the unmodified *BEB*), and that somebody is in breach. From this list it can be inferred that that person is blameworthy. But whilst a monotonic logic allows the inference to be drawn no matter what premisses we add to the list, in a ‘nonmonotonic’ logic the inference is retractable depending on the premisses we add. For example, if we add that the person also has an excuse, then in a nonmonotonic logic this new information may allow the original inference to be retracted, such that we may no longer conclude that the person is blameworthy. On the nonmonotonic approach, the *ceteris paribus* clause makes the inference from breach to blameworthiness defeasible by encoding this information-dependent retractability.

An initial problem with this attempt to salvage the overall reading of *BEB* is that it faces yet another counterexample, this time from justified breach of *overall* obligation.

Consider, for example, that we sometimes face dilemmas between incomparable and mutually unsatisfiable obligations. A famous example is owed to Sartre (2007: 30), or at least to a student of his during World War II, who faced “the choice of going to England to join the Free French Forces—which would mean abandoning his mother—or remaining by her side to help her go on with her life.” This is clearly a dilemma between incomparable and mutually unsatisfiable obligations. By their incomparability neither obligation defeats the other, and so both are undefeated in the circumstances, making them facts at the overall level. This means it would be justifiable for him to act in satisfaction of either obligation, so that either obligation is justifiably breached in the course of satisfying the other. Thus, if he acts in satisfaction of either obligation *for* that reason—if he is justified in breaching the other obligation—then he cannot, as we said before, be blameworthy for that. We therefore have a straightforward counterexample to the nonmonotonic *ceteris paribus* version of *BEB* on its overall reading.

Perhaps the proponent of *BEB* will reply that this approach can accommodate the new counterexample, by making other things unequal not only when excused but also when justified. At this point, the modified *BEB* may seem so far removed from the original that its proponents are probably fairly accused of clutching at straws. But even if not, the nonmonotonic *ceteris paribus* approach still faces the following dialectical problem. In the context of the MORALISTIC ARGUMENT, *BEB* is a conditional proposition intended to hold as a matter of conceptual necessity. But nonmonotonic conditionals are not like this: their inferential potential depends on the informational context. If *BEB* is context-variant, however, that means the MORALISTIC ARGUMENT only goes through when the context supports it. Whenever it does not—if there is an excuse (or perhaps a justification)—the MORALISTIC ARGUMENT finds no contradiction and thus sits happily alongside the possibility of tragedy. So, any nonmonotonic version of the *ceteris paribus* approach threatens to bring moralism so close to anti-moralism that it becomes puzzling what the debate was all about.

This concludes my case against *BEB*. We can summarise as follows. On a first pass attempt at giving ‘obligation’ in *BEB* a suitable overall or contributory reading, it faces counterexamples from excused and justified breach respectively. The proponent of *BEB* may respond by adopting either a monotonic or nonmonotonic version of the *ceteris paribus* strategy in a bid to salvage the overall reading. Both versions deal successfully with excuses, but they each face further problems of their own. The monotonic version

invalidates the MORALISTIC ARGUMENT, making the various positions on the debate over moralism ill-defined. The nonmonotonic version faces a counterexample from justified breach of overall obligation, and moreover threatens to collapse the dialectical distance between moralism and anti-moralism entirely.

In this section, I argued that we should not accept *BEB*. What follows from this for the debate over moralism? Firstly, it follows that the MORALISTIC ARGUMENT should be rejected, and with it, the moralist position on the debate. Secondly, in tandem with the section 2.3 defence of *BEF*, it follows that our anti-moralism should take a ‘soft’ rather than a ‘hard’ form.

2.5. Interpreting Feelings of Responsibility

Over the last two sections, I argued that we are not forced to choose between the equally unpalatable options of moralism and hard anti-moralism. To resolve the moral luck dilemma as it arises for tragic cases, we may simply soften our anti-moralism. But why, then, have so many anti-moralists preferred the hard route? In this final section, I diagnose just why the hard route has looked so attractive to those with anti-moralist sympathies. To anticipate, I suggest that it has to do with a certain paucity in our moral-psychological repertoire, one that must be rectified if anti-moralism is to be at all viable.

The germ of anti-moralism, quite generally, lies in the powerful phenomenology of tragedy. It is all too familiar for our faultless failures to leave us feeling responsible, and at least sometimes, it strikes us as apt to feel this way. This is what the Dolphin House tenant so poignantly conveys, in saying “I must be a terrible parent to leave my children living in these conditions.” What pretheoretically unites anti-moralists of all stripes is that they, like the tenant, see it as a fundamental datum that tragic situations can make us apt to feel responsible. The force that anti-moralists ascribe to this phenomenology is plain to see. Consider Marcus’s remarks on the situation of Sartre’s student, whose moral dilemma leaves him quite faultless in failing to satisfy whichever obligation he opts against. She says (1980: 131),

it is inadequate to insist that feelings of guilt about the rejected alternative are mistaken and that assumption of guilt is inappropriate. Nor is it puritanical zeal

which insists on the reality of dilemmas and the appropriateness of the attendant feelings. For dilemmas, when they occur, are data of a kind.

What Marcus calls “guilt” plays the role of a feeling of responsibility, which she says is not “inappropriate,” meaning that Sartre’s student is quite apt to feel it in response to his failure with respect to the “rejected alternative.” Conversely, the moralist must regard these feelings as pathological: as Marcus puts it (1980: 131-2), “feelings of guilt or pangs of conscience are viewed as, at best, sentimental.”

Notice, however, that the phenomenology leaves latitude as to the name and nature of the relevant feelings. All it requires is, firstly, that they are feelings of *responsibility*—some kind of *negative retrospective* feeling about *what we did*—and secondly, that those feelings are apt. And it is here that anti-moralists begin to diverge. Though we all agree about the force of the phenomenology, we disagree about how to interpret the feelings of responsibility involved. We can see from the above passage that hard anti-moralists like Marcus interpret them as forms of blame, such as feelings of remorse or guilt. This, I believe, is the source of the hard anti-moralists’ error, and explains why they endorse *BEB* over *BEF*. The explanation is simple enough. If we interpret the feelings as blame, then given the faultlessness of tragic situations, we cannot accept that blameworthiness entails fault: for blame is exactly what we identify as fitting in those situations. So, for hard anti-moralists, accepting *BEF* would *undermine* the force of the phenomenology, recognising which is the beating heart of anti-moralism as such.

This shows that despite the hard anti-moralist’s staunch opposition to the moralist, they both agree on something deep: they both interpret the phenomenology in terms of forms of blame. The disagreement between them lies downstream of this interpretation: the moralist claims that blame in tragic cases is pathological, thus denying the force of the phenomenology. But for soft anti-moralists like me, this squabble is premised on a shared misinterpretation of the phenomenology. In interpreting the relevant feelings as blame, the moralist and the hard anti-moralist in fact stand united against the soft anti-moralist, even as the anti-moralists stand together against the moralist in accepting the force of the phenomenology.

The resulting three-way standoff is a delicate affair indeed. We begin to get a handle on it by reconstructing the MORALISTIC ARGUMENT in a way that does not build in any prejudicing commitments about how to interpret our feelings of responsibility.

NON-COMMITTAL MORALISTIC ARGUMENT

- (1) Situations of faultless failure are possible, where people both
 - (1a) Stand in breach of an obligation,
 - (1b) But are not at fault in so doing.
- {assume for reductio}*
- (2') Anyone who breaches an obligation thereby fittingly experiences feelings of responsibility, Ψ , for so doing.
- (3') No one fittingly experiences feelings of responsibility, Ψ , for doing anything without being at fault in so doing.
- (4) A person, who faultlessly fails, stands in breach of an obligation.
{from 1a}
- (5) The same person does not stand in breach of that very obligation.
{from 1b, 2', 3'}
- (6) Since (4) and (5) are contradictory, (1) must be false.
{from 2', 3', 4, 5}

We have here a valid argument schema, the instances of which vary in soundness depending on how the relevant feelings of responsibility get interpreted. The moralist substitutes blame into Ψ and takes the resultant argument, the MORALISTIC ARGUMENT, to be a *reductio* of the possibility of tragedy. As we saw in section 2.4 this is mistaken: substituting blame into Ψ renders the argument unsound at premiss (2'). The hard anti-moralist joins with the moralist in plugging blame into Ψ , but, impressed by the force of the phenomenology, doubles down on the possibility of tragedy and is led to locate the unsoundness at (3'). This too is a mistake, as we saw in section 2.3, since it is simply not plausible to deny that fault is required for blameworthiness. On my preferred soft anti-moralist alternative, it is more promising to plug some feeling of responsibility besides blame into Ψ , locating the unsoundness at (3') rather than (2'), such that breach suffices for the feeling to be apt, whilst fault is not necessary for it to be.

Seen this way, the surprising abstractness of pretheoretical anti-moralism is starkly visible. Whether 'hard' or 'soft,' we agree that however we interpret the relevant feelings, the unsoundness should be located at premiss (3'). Whatever the relevant feelings are,

we agree on their indifference to fault. So, despite the infighting amongst anti-moralists over how to interpret the phenomenology, we must not forget that our agreement about its force reflects an ethical intuition whose content is highly abstract, and which cuts right to the heart of pretheoretical moral experience. We anti-moralists believe that in some situations, we are doomed to fail in ways that make us apt to feel responsible, no matter how faultlessly we may have conducted ourselves. Of course, as we have seen, the ‘hard’ amongst us go on to join with the moralists in how they interpret the relevant feelings, and so cannot but be led to dodgy views about the conceptual relations between blameworthiness, obligation and fault. But what they share with the moralists is quite superficial compared to the abstract intuition they share in with their softer anti-moralist fellows.

I believe that the considerations advanced in this chapter are cumulatively sufficient to show that hard anti-moralists are wrong to interpret the phenomenology in terms of blame. But they can hardly be blamed for doing so. What other feelings of responsibility could they plausibly have reached for? Our canonical catalogue of moral emotions just does not seem to deliver. It is notable what Marcus says about regret in this connection. Of the aftermath of choice under tragic dilemmas, she says that “to describe our feelings about the rejected alternative as ‘regret’ seems inadequate” (1980: 130n9), and that “To insist that ‘regret’ is appropriate rather than ‘guilt’ or ‘remorse’ is false to the facts” (1980: 133n12). We find the same point in Marcia Baron, who writes that agent-regret “seems out of place [...] because it is so mild,” and that whatever the relevant feeling, it is “more than regret” (1988: 261). In other words, although agent-regret may indeed be a feeling of responsibility, it is not, or not sufficiently, possessed of a ‘moral’ quality that would do justice to the phenomenology as it strikes us.

Agent-regret is just so *vanilla*. Other things equal it can befit a failure to meet just about any reason for action, very trivial ones included. It is the kind of feeling that is appropriate to the bus we just missed (precipitating a twelve-minute wait), or to the unordered tiramisu that now looks so delicious, and so on. We may want to say that the relevant ‘moral’ quality attends the failure to satisfy very *weighty* reasons, but that will not do. Weight is a quantitative notion, but we are after a qualitative difference. No matter how long the wait for the next bus, or how big or delicious the tiramisu, our agent-regret will intuitively never amount to what the tenant of Dolphin House feels. Or

at least, if it can, the explanation will not lie in the weight of the reasons left unsatisfied.³¹ Whatever the feeling is, it must reach qualitatively beyond mere agent-regret. As far as painful feelings with a special ‘moral’ quality go, forms of blame like guilt and remorse would appear to fit the bill. If that is all that canonical moral psychology has to offer, this leaves the anti-moralist with little choice but to go ‘hard’ or go home.

2.6. Concluding Remarks

Tragedy is an everyday phenomenon: sometimes it seems that no matter how faultlessly we conduct ourselves, we are doomed to fail in our obligations, making us apt to feel responsible as a result. Sensing that this is unfair, moralists about obligation deny the very possibility of tragedy. Somewhat ironically, this makes them liable to pathologise our everyday experience, or, worse still, to vindictively find fault. If we are to resist these moralistic tendencies, we must reject the MORALISTIC ARGUMENT. I have argued that we should be careful not to do so in a way that falls prey to the moral luck dilemma as it arises for tragic cases. For unless we can ‘soften’ our approach, we are led to a ‘hard’ anti-moralism that is no less objectionable than the moralism it seeks to oppose.

I suggested that to avoid this mistake, it is incumbent on us to extend our repertoire of ‘moral’ emotions. In particular, I suggested that we need to account for a novel feeling of responsibility that is (a) of sufficiently ‘moral’ quality to justice to the phenomenology of tragedy, and (b) apt in the absence of fault. Developing such an account is the goal of Part Two. To anticipate, I believe we should understand the relevant feelings as *ordinary agent-regrets in response to failures to satisfy extraordinary reasons for action*. My account, if plausible, will go some way to vindicating soft anti-moralism as a way out of the moral luck dilemma for tragic cases. Absent some such account, I have argued, we cannot but struggle to avoid responding felicitously to the victims of tragedy, like the tenant of Dolphin House.

³¹ See the preamble to Part Two for lengthier consideration of this point.

Part Two

Tragedy

The contingency of the world can frustrate whatever control we may have had over how our lives pan out. Sometimes it even leaves us apt to feel responsible for what was not in our control, what was not our fault, lending a certain tragedy to those situations. Tragedy is especially troubling because we struggle, in ethical thought, to respond to it felicitously. We may insist on the Fault Principle on blameworthiness: a moralistic denial of tragedy that forces us to pathologise its victims, or worse still, to find fault in their conduct. Recoiling from this, we may be led instead to a hard anti-moralist denial of the Fault Principle, taking the victims of tragedy to be blameworthy despite their lack of fault.

Part One made the case that we should reject moralism, but that our anti-moralism need not be 'hard.' We saw, however, that soft anti-moralism must appeal to a feeling of responsibility that stands somewhere between agent-regret and blame. The feeling must, like agent-regret, be capable of aptness in the absence of fault, whilst also, like blame, be of a 'moral' quality sufficient to do justice to tragedy. I will now present the outlines of

an account, which it is the overall aim of Part Two to defend. In a nutshell, I will claim that the feelings of responsibility at issue are *ordinary agent-regrets over failures to satisfy extraordinary reasons*. I shall begin with some general thoughts about value, reasons for action, and what goes on when we fail to act accordingly.

When an action would be good to do, that is due to its character as a response to the things that matter. What exactly they make it good to do at a given time depends on the circumstances. That it would be good if we were to do something at a given time explains why we have a reason to do it at that time.¹ Exactly when our reasons say to act can be more or less precise, again depending on the circumstances. In general, when an agent *S* has a reason *R* to do ϕ , *S* only fully satisfies *R* if *S* does ϕ within some window of time *t*.² Reasons for action are thus indexed to a window of time within which they are fully satisfiable, that is, reasons to ϕ are always strictly speaking reasons to ϕ -at-*t*. That window may be so short that *S* fails to fully satisfy *R* unless *S* ϕ s at a very precise point in time, or so long that *S* fully satisfies *R* should *S* ϕ at some point in their life. But there will always be *some* window—however short or long—if only because our lives are of a finite duration. In general, then, reasons have a time for action built into them, and that time can always come to pass.

Once that time has passed, the good there would have been in acting is, to an extent, lost. But only to an extent. In most cases, there is often something next best that can still be done. Whether there is anything next best left to do, and what it would consist in, is some function of what was at stake in acting in the first place, combined with the present circumstances. So, if the time has now passed to do something it would have been good to do, then what made it so, plus what can now be done, determines what would be next best to do—if anything at all. This explains why, once we fail to fully satisfy a reason for action, it often becomes, without further ado, a reason to do the next best thing.³ Call

¹ See Chapters Three and Four for more on how our reasons for action are explained by the value of what they say to do.

² Note that for *S* to fully satisfy *R*, *S* need only do ϕ , rather than doing ϕ for *R*. Sometimes, *S* may have a reason, *R*, to ϕ , where ϕ is such that *S* cannot do it except by doing it for *R*. It remains generally true, nonetheless, that *S* fully satisfies *R* only by doing ϕ . For more discussion of these special cases, see Chapter Five.

³ This is an implication of what Joseph Raz (2004: 189) calls the ‘Conformity Principle’: “One should conform to reason completely, insofar as one can. If one cannot, one should come as close to complete conformity as possible.”

these *reasons of repair*: to make up, as far as is now possible, for the loss of whatever value there would have been in doing as our original reasons said to do.

In general, when events transpire that cause a loss of value, it is, to that extent, fitting to feel bad in response. For example, if, during a storm, a child is injured by a falling tree, it would be fitting to experience some sort of negative retrospective feelings towards this event. What should we call feelings like this? I follow Bernard Williams (1976) in using the word ‘regret,’ though there is, of course, a certain arbitrariness about this. Others in the literature prefer ‘sorrow,’ ‘lamentation’ or ‘dismay,’ and these words all sound fine to me too. The important thing is not to let our linguistic practices surrounding the feeling at issue get in the way of the feeling itself. Here, as everywhere, our interest lies in the phenomenon—what we call it does not matter so much.

When it would have been good for us to act, but the time to do so comes and goes, value is certainly lost, and regret is fitting to that extent. But to say ‘value is lost’ does not seem to capture the whole truth. For not only is value lost, but moreover: *we lost it*. The events did not so much unfold around us, as *through us*, that is, through what we did or omitted to do. In cases like this, it seems fitting that our regret qualitatively reflects our agential role in the loss of value. Again, following Williams, let us call this type of feeling *agent-regret*. In his words (1976: 123-4),

there is a particularly important species of regret, which I shall call ‘agent-regret,’ which a person can feel only towards his past actions (or, at most, actions in which he regards himself as a participant). [...] [S]entiments of agent-regret are different from regret in general, such as might be felt by a spectator, and are acknowledged in our practice as being different. The lorry driver who, through no fault of his, runs over a child, will feel differently from any spectator, even a spectator next to him in the cab [...]. We feel sorry for the driver, but that sentiment co-exists with, indeed presupposes, that there is something special about his relation to this happening, something which cannot merely be eliminated by the consideration that it was not his fault.

Williams is suggesting that whilst a spectator can regret the event as something that merely happened, akin to a falling tree, the lorry driver has no such refuge. For the lorry driver stands in a different relation to the event: they not only bear witness to a loss of

value, but *they were the one who lost it*. And this seems to explain why they are vulnerable to a quality of negative retrospective feeling that the spectator is not: one expressible, says Williams (1976: 124), not just in condolences, but in “some kind of recompense or restitution.” So, unlike mere regret, it makes sense to think of agent-regret as a feeling of responsibility, a feeling Williams claims (1976: 125) “it would be a kind of insanity never to experience.” But crucially, unlike other feelings of responsibility—in particular, forms of blame such as remorse—it may be apt in the absence of fault.⁴ What makes it apt is the occasion of a loss of value we occasioned, which is possible even if we are, like the lorry driver, quite faultless in the situation.

Before moving on, some remarks on my conception of agent-regret. Firstly, on some views, agent-regret is a ‘moral’ emotion, in the sense that, as Marcia Baron (1988: 262) puts it, it “is felt toward the sorts of things which, if done deliberately, would properly occasion guilt.” This does not follow on my view, since we may occasion a loss of value, and thus be apt to feel agent-regret, without the loss being so serious as to make us aptly guilty had we occasioned it deliberately. For instance, someone could spill tomato soup down their nice shirt and aptly feel a pang of agent-regret. Secondly, a natural thought is that there is no point feeling responsible for outcomes not of our choosing, and that agent-regret must therefore exclusively concern the choices we made. If we take this view, however, we struggle to accommodate agent-regret for mistakes, accidents, inadvertence, and so on. We typically do not choose to spill soup down ourselves, but we surely can feel agent-regret for having done so. My view allows for this: in making agent-regret apt to a loss of value we occasioned, its basic object is what we did, not what we chose to do. Thirdly, my view makes apt agent-regret a *pro tanto* normative phenomenon, in that, for anything we did, we fittingly feel agent-regret for any loss of value thereby occasioned. If what we did lost value in more than one respect, we fittingly feel agent-regret for the value lost in each respect. And if what we did happens also to realise value in other

⁴ This way of distinguishing agent-regret from remorse is from Williams (1976: 126), and it has been endorsed and expanded upon by Marcia Baron (1988), R. Jay Wallace (2013) and David Sussman (2018). Like these authors, I take the phenomenology to support this distinction, but see Daniel Jacobson (2013) for opposition. Jacobson argues, in effect, that the phenomenology that interests Williams can be explained away: the seemingly fitting feeling we introduce agent-regret to explain may in fact be admirable, though unfitting, remorse. Though I cannot reply to Jacobson here, Chapter 3 offers an argument for the fittingness of agent-regret independent of the phenomenology in dispute.

respects, we would still, on my view, fittingly feel agent-regret for the value we lost in whatever respects we did—even if, overall, we realised more value than we lost. Fourthly, and finally, the fact that we can make such overall assessments does not imply the existence of some single ‘on-balance’ value in what we did, and I do not see what would be gained from positing its existence. I assume, therefore, that there exists no such ‘on-balance’ value—only value in different respects. This is why I neither recognise nor consider the phenomenon R. Jay Wallace (2013: 51) calls “all-in regret.”⁵

Taking these remarks as read, let me develop my view of agent-regret. Agent-regret befits the occasion of a loss of value we occasioned, so once the time has come and gone to fully satisfy a reason for action, we are apt to feel agent-regret. But, as mentioned, all is not usually lost; there is often something next best left to do, which makes up for the loss to some extent. I follow John Gardner’s (2018a: 140) suggestion that agent-regret is “an apt response to what has so far gone unrepaired.” In the moment that the time has come and gone to fully satisfy a reason, we have yet to repair the whole loss of value, so we are apt to feel agent-regret for the full extent of that loss. But if we then do the next best thing, we are thereby apt to feel less agent-regret, for we leave less so far unrepaired: we make up for the value there would have been in doing as our original reason said to do. Of course, since, by hypothesis, the time for *that* has been and gone, we can only make up for it as far as is now possible. Doing what is next best is never quite the same as fully satisfying the original reason: there will always be something left undone, the value of which the passage of time has made inaccessible to us. “There must,” as Gardner says (2018a: 142), “always linger some residual ground for agent-regret, some unrepaired and irreparable remainder.”⁶

⁵ Wallace (ibid.) believes regret can be “all-in” because, on his view, regret in general involves the retrospective preference that things should be otherwise, and those preferences can be ‘on-balance.’ But on my view, though agent-regret may involve retrospective preference, what makes it fitting is not the preference but the loss of value we occasioned. So even if it is true, as Wallace suggests, that preferences for what could have been can be ‘on-balance,’ the value of what could have been cannot—and so I recognise no normative phenomenon of ‘all-in’ agent-regret.

⁶ It may seem that for some reasons for action, no irreparable remainder is left when we fail to satisfy them. For instance, if I spill my cup of tea, I can simply brew another. Or if I lose your £10 note, I can simply give you another. Is there really an irreparable remainder even in cases like these? I say: ‘yes,’ if there is a reason left unsatisfied. It may seem otherwise, I believe, because the reasons we tend to imagine unsatisfied in such cases specify a window of time for action so long as to make it almost immaterial when the action is done. Of course, if only *this* cup of tea

Repair can be, but need not be, owed to others. Ari, who has a reputational reason to look presentable, has a reason to take reparative measures if he spills tomato soup all down himself: perhaps to clean, change or cover up his shirt (with a jumper, say), as the case may be. But equally, Brett, who spilled soup over his colleague Cindy, may also have a reparative reason, only this time owed to Cindy—to clean her shirt, find a replacement, or lend his jumper to cover up the mess, as the case may be. For both Ari and Brett, there are the irreparable remains, that is, what would have been good about not ruining the shirt in the first place, which no amount of cleaning, changing or covering up can quite make up for. But of what is repairable, both Ari and Brett can repair what has so far gone unrepaired by doing these things. The difference in what would have been good in what Ari and Brett originally had reason to do explains why Brett, but not Ari, owes his repair to another, in this case Cindy. In spilling soup over Cindy, Brett threatens to damage her reputation, which a good colleague would not do (or not unnecessarily). But Brett's bond of collegiality—unlike Ari's reputation—matters as a constituent not just of his own life, but Cindy's life too. So, in spilling the soup, what Brett did represents a loss not only for himself, but for Cindy, in a way it simply does not for Ari. This is why Brett's reparative actions make it up to Cindy, whilst Ari's doing the same only makes it up to himself.

In spilling the soup, both Ari and Brett leave an irreparable remainder, which makes them apt to feel agent-regret whether or not they repair later on. But though their agent-regret befits their failure insofar as they leave it unrepaired, it is not fitting in virtue of being repair. Simply feeling agent-regret cannot on its own be the next best way to realise the values at stake in not spilling the soup to start with because feeling, unlike action, is immaterial. If there are reparative reasons to feel agent-regret, that can only be because it contributes to doing something next best. This is typically the case: since agent-regret befits what has so far gone unrepaired, its painful quality typically serves to compel the repair that would ease the pain. And when it is our relationships we have damaged, there is often no repair without reconciliation, and no reconciliation without an expression of

will do (I am on a short break and have no time to brew another), or if you need *this* £10 note right now and I can only give you another tomorrow, the irreparable remainder reasserts itself. But if what would be good about providing an alternative now is literally no different from what would have been good about not losing the original back then, then all this shows is that, in fact, I had no particular reason not to lose the original. In that case, I leave no irreparable remainder, but nor did I flout any particular reason.

agent-regret—typically through an apology—which is, of course, easier to do effectively when agent-regret is actually felt. But there need be no such reparative reasons to feel agent-regret for us to be apt to feel it. To repeat: what makes it fitting is just the occasion of the loss of value we occasioned, to the extent that we have yet to make up for it.

We said, however, that once we fail to fully satisfy our reasons, there will always be an irreparable remainder that makes us apt to feel agent-regret. Since, by definition, we cannot repair the irreparable, does that not make us apt to feel agent-regret over it for evermore? That already sounds like a worrying implication. The worry compounds itself when we see that, due to value pluralism, in satisfying one reason we always leave some other reason unsatisfied as a result. Over time, will we not leave so many reasons unsatisfied that our inner lives are apt to be consumed with misery? I think these worries are overstated. To start with, our doing the next best thing may almost fully make up for what we did, leaving us apt to feel only trifling agent-regrets. Even if that is not so—or indeed even if we fail to repair at all—our reasons are often themselves trifling, in which case the full measure of agent-regret we aptly feel is itself trifling. I am not sure why we should think that trifling agent-regrets need bother us at all: perhaps our inner lives are awash with myriad feelings, many of which never rise to conscious salience. Finally, and most importantly, I believe that even where our agent-regret rises to phenomenological salience, the normative sway of the reasons that make it fitting can be counteracted in two ways. Many reasons for action depend for their existence on conative attitudes—our goals, plans, desires, and the like—such that we can escape these reasons by simply revising the relevant conative attitudes. Moreover, even if we cannot or do not escape our reasons for action in this way, we may have reasons of psychological convenience which defeat the reasons making it apt for us to keep feeling agent-regret. In other words, it often makes sense to move on from our misery by *letting go* of what makes it fitting, or by *repressing* the feelings themselves.

Not all reasons for action are like this, however. Some reasons for action seem not to depend for their existence on our conative attitudes, and this attitude-independence makes them, in this sense, *inescapable* at will.⁷ And some reasons for action seem not to

⁷ The sense in which reasons are ‘inescapable’ at will, when they are, is discussed in detail in Chapter Four. To forestall possible confusion, suffice it to say, for now, that reasons that seem ‘inescapable’ in the sense at issue need not seem indefeasible. Our having a certain reason may

be defeasible by countervailing reasons of convenience, however weighty they may be, because they are normatively structured to exclude such considerations from guiding our conduct. Such reasons have an exclusionary force, in virtue of which they strike us as *requirements* to do as they say, even if, in some cases, what they say to do is unsupported by the balance of reasons.⁸ Inescapability and requiring force seem to be logically independent properties of reasons. Some reasons are apparently inescapable but lack requiring force, for example, we all have reasons to give to charity whether or not we are attitudinally disposed to do so, though we think of charitable giving as, by definition, going beyond requirements. Conversely, some requirements seem escapable at will, as when our personal dietary regime excludes getting a takeaway tonight however delicious, cheap or convenient it may be, though it ceases to be a reason for us not to act for those reasons if (say) we stop caring about looking slim. But equally, some reasons seem to be both inescapable and requiring, and our *obligations* have sometimes been identified with such reasons.⁹ It is characteristic of obligations that they are not escapable by simply revising our conative attitudes, and that they can require us to act even if it is inconvenient to do so, or otherwise contrary to the balance of reasons.

Suppose that this account of obligation is correct. If so, I would be presumptively justified in taking what I said about repair and agent-regret to apply, *mutatis mutandis*, to obligations left unsatisfied. Just like ordinary reasons, when we flout our obligations—or, as we usually say, when we are in breach—they call for repair as far as is now possible, while making us apt to feel agent-regret unless and until that happens. Presumably they do this with the same normative force with which they agitated for their own satisfaction in the first place. That is, since reasons left unsatisfied generally become reasons to repair,

seem impervious to changes in our conative attitudes whilst being so trifling in weight as to be easily defeasible by other reasons of ours, even reasons that strike us as attitude-dependent.

⁸ I defend the ‘exclusionary’ character of some reasons for action, and their ability to explain how some reasons are normatively structured as requirements, in Chapter Five.

⁹ For example, Raz (1994: 40) seems to do so when he says, “Duties are but a special kind of categorical reasons, they are peremptory reasons, reasons which exclude consideration of an prevail over certain categories of reasons.” Many subsequent authors, such as John Gardner and Timothy Macklem (2004: 465), see themselves as followers of Raz in accepting this view of obligation. I hesitate to say that this was Raz’s considered view, for reasons brought out in Samuel Scheffler (2004: 261-3), and which I discuss in Chapter Four. I prefer to say that the proposed view of obligation is very much inspired by Raz, though perhaps not Raz’s own.

obligations left unsatisfied become obligations to repair, which presumably require us to repair even if greatly inconvenient, and which we cannot easily escape. And to the extent that flouted obligations go unrepaired, they presumably make us apt to feel agent-regret with the same force, that is: we cannot cease to make it apt by escaping the reason that makes it so, nor can we let our psychological convenience guide any resolution to stop feeling that way. I am suggesting, then, that when reasons left unsatisfied have obligatory force, it is difficult to justifiably let go of or to repress the agent-regrets we are apt to feel. Or, as Gardner puts it (2005: 105), “Wrongful action [...] leaves us with regrets that are hard to expunge and the repression of which is hard to justify.”¹⁰

On the view I propose, of course, there is always an irreparable remainder when we leave our obligations unsatisfied, just as with ordinary reasons for action. Given what I said about the associated agent-regrets, does this not just reintroduce the worry that it is therefore fitting to feel them forevermore? Well, yes, it does—and that is the whole point. The worry is that when we feel ordinary agent-regret in response to extraordinary reasons left unsatisfied, those feelings are difficult to justifiably let go of or repress: they threaten to stay with us, a misery we are forever apt to feel. It is precisely this worrying feature that I suggest lends agent-regret the special ‘moral’ quality we have been looking for. It is this, I claim, that makes such agent-regret suitable to meet the soft anti-moralist desideratum from Part One: a feeling of responsibility capable of aptness in the absence of fault, whilst being of a sufficiently ‘moral’ quality to do justice to tragedy. My thought is that if we can help ourselves to this account, we can rescue ethical thought from the excesses of both moralism and hard anti-moralism.

Certainly, however, we cannot *simply* help ourselves to the account. There are many possible objections to it, some more urgent than others. Let me briefly address the less urgent amongst them. First of all, we could accept the explanation for why some fitting agent-regrets are hard to move on from, but still remain worried about the idea that they could be fitting forevermore. Note, however, that to meet the anti-moralist desideratum, the relevant fitting agent-regrets need only be difficult to move on from—not impossible. I agree that forever fitting feelings would be undesirable, though it is notoriously difficult

¹⁰ Gardner develops this view in more detail at (2005: 102-5), and the same view is expressed in earlier work with Macklem (2004: 467-8; 467n36), though it is hard to say whether Gardner remained committed to the view in his later work.

to explain why they are not forever fitting.¹¹ I do think that agent-regrets made fitting by failing to satisfy attachment-dependent reasons can be escaped by extricating ourselves from the valuable attachments they depend on, though this, of course, should worry us no less: we escape our fitting misery only by destroying the constituents of the lives we are leading.¹² The real worry, then, is that we are left with no possibility of redemption. I must admit, like many others, to having no clear idea of how we redeem ourselves, other than that it cannot proceed by destroying our valuable attachments: we somehow have to fold the grounds of our misery into a living story that makes sense of it. I can only plead that the worry, if unanswered, has ramifications in ethics that extend beyond my own account. It is no more incumbent on me to address this worry than anyone else.

Another objection targets the alleged ‘moral’ quality of the relevant agent-regrets, which is, in a nutshell, that they justifiably plague us. The objection is this. Obligations, like any reasons, can be trifling. For instance, Daphne may forget that she promised Ellis to meet for a quick coffee over lunch. In all likelihood, Daphne’s promissory obligation to Ellis is not so important, so the agent-regret she is apt to feel will be correspondingly trifling. But my account seems to imply that Daphne’s feelings nonetheless possess the relevant ‘moral’ quality. The objection is: is it not hyperbole to describe these feelings as ‘plaguing’ Daphne? That would, I agree, be hyperbolic—but I am not committed to describing her feelings in that way. For I can claim that agent-regret is apt to plague us only in response to breaches of *important* obligations. To possess the relevant ‘moral’ quality, our agent-regret must not only stay with us, but it must also be sufficiently grave.

But now, if agent-regret lacks a ‘moral’ quality unless it is sufficiently grave, it may begin to look questionable whether the difficulty of justifiably shaking it off really plays any role in the explanation. To rebut this, I need to provide a case that clearly shows that the sheer gravity of agent-regret is not alone sufficient to secure its ‘moral’ quality. The dimension of gravity is tricky to isolate because, very often, the more intuitively weighty the reason, the more plausible it is to attribute the intuition to the reason in fact being

¹¹ See Gardner (2018a: 127-9) for his preferred explanation, though it remains inchoate. More developed explanations are available in Agnes Callard (2017) and Oded Na’aman (2021). See Christopher Howard (2022) for a recent argument against Callard and Na’aman, which builds on ideas in Dan Moller (2007). For an expression of both theoretical and practical exasperation at our inability to give an adequate explanation, see Berislav Marušić (2018).

¹² See Chapter Four for extended discussion of this possibility.

an obligation. Bearing that in mind, consider the following sort of case. Suppose that blueberries are your favourite fruit. Every day you eat blueberries makes that day go that bit better. You love blueberries so much that the life improvement from every day that you eat them is not subject to diminishing returns, even if, in the course of any day, the improvement to that day from each one you eat is so subject. If you could get a day's fill of blueberries by a minimal action, no strings attached, the improvement to that day would be a trifling reason to do it. To your delight, you happen to win a lottery for a daily delivery of blueberries for the rest of your life. To collect your prize, you just have to submit the winning ticket by next week. Presumably, the prospect of an improvement (however trifling) to each day of the rest of your life adds up to an extremely weighty reason to submit the ticket. Thus, if you forget to submit the ticket in time, you are apt to feel extremely grave agent-regret. But however grave, it seems implausible that your feelings thereby possess a 'moral' quality. You may be right to kick yourself about the blueberries, but there would surely be nothing amiss if you were to get over it before long, and probably something amiss if you did not.

This concludes my discussion of what I believe are the less urgent objections to my account. There are, however, several more serious objections that merit fuller discussion. The account I have presented can be thought of as having basically two parts: the claim that agent-regret generally befits reasons for action left unsatisfied (insofar as they have so far gone unrepaired), plus the claim that when those reasons are important obligations, the agent-regret we fittingly feel has a special 'moral' quality. I have in mind objections targeting each part of my account. Firstly, we may agree that what we did may call for repair, whilst remaining skeptical that we are thereby apt to feel painful feelings of agent-regret. Unless I can rebut this skepticism, my account may seem to express little more than an idiosyncratic preference for self-flagellation. Secondly, we may accept my view of ordinary agent-regret, but remain skeptical about the extraordinary reasons that I claim imbue the relevant agent-regrets with a special 'moral' quality. Our doubts may target either component of my analysis of obligations, that is, their status as inescapable reasons, requiring reasons, or indeed both.

Part Two aims to address these objections. Chapter Three, 'Why Feel Agent Regret?', offers an argument as to why we are apt—and not just masochistic—to feel agent-regret when we fail to satisfy reasons for action. Chapters Four and Five move on to the theory of obligations. Chapter Four, 'On the Reasons We Cannot Escape,' attempts to advance

our understanding of the sense in which our reasons are ‘inescapable,’ when they are, by exploring why we have attachment-dependent obligations, when we do. Chapter Five, ‘In Defence of Second-Order Reasons,’ defends the concept of a second-order reason—essential to explaining the requiring force of obligations—against a recent argument for its incoherence. Thus, Part Two is a sustained attempt to defend the proposed account of a feeling of responsibility that satisfies the soft anti-moralist desideratum. Although there will no doubt remain stones left unturned, I hope that my efforts in Part Two will suffice (a) to lend credibility to the account, and (b) to entitle me to it in Part Three.

CHAPTER THREE

Why Feel Agent-Regret?

3.1. Agent-Regret Eliminativism

As indicated in the preamble, my aim in Part Two is to vindicate an account of a feeling of responsibility fit to satisfy the soft anti-moralist desideratum explained in Part One. The relevant feeling must (a) be fitting in the absence of fault, and (b) be of a sufficiently ‘moral’ quality to do justice to cases of tragedy. In a nutshell, I claim that we should think of these feelings as ordinary agent-regrets in response to failures to satisfy extraordinary reasons for action. Chapter Three aims to address a deep objection to the claim that it is fitting to feel agent-regret for failing to satisfy our reasons for action *at all*, extraordinary or otherwise.

Let me briefly recap the view of agent-regret presented in the preamble. Sometimes, events transpire that cause value to be lost. Such events are in this sense ‘bad.’ When bad things happen, it is, I take it, generally fitting to feel bad in response. For instance, if we witness someone being hit and injured by a falling tree, we would be apt, I think, to feel regretful about this without further ado. But sometimes when bad things happen, they happen to be bad things that we are responsible for. For instance, we may, like Bernard Williams’s (1976: 124) lorry driver, we may hit and injure someone through no fault of our own. As with the falling tree, a person ends up getting hit and injured. But we bear a different relation to the event: not only does the bad thing happen, but it is moreover

something we did. To merely feel regret does not seem enough—we seem apt moreover to feel an *agent's* regret for the bad thing *we did*, at least to the extent that it has so far gone unrepaired.¹

No doubt the painful feeling of agent-regret will be all too familiar. But does it really make any sense that we suffer such feelings? On closer scrutiny, the fittingness of agent-regret can begin to look rather doubtful. This is not just because, as the everyday wisdom has it, ‘There is no use crying over spilt milk.’ There can, of course, be good pragmatic reasons to repress fitting feelings such that, all told, in the circumstances, we should not feel them. But at least since Justin D’Arms and Daniel Jacobson (2000), it is usually thought, in my view rightly, that what makes feelings apt in the first place is independent of extrinsic, pragmatic concerns. We do not cry to serve a function beyond the situation at hand, but rather, as Amia Srinivasan (2018: 132) would put it, to properly *appreciate* the badness of what we did.² Unfortunately for these views, it never gets explained quite why proper appreciation has to involve painful feelings at all, rather than, for instance, emotionless evaluative judgments to the effect that we did a bad thing. If that is possible for us, why should we not just *do away* with the needless pain of agent-regret entirely? I call those who answer ‘yes’ *eliminativists* about agent-regret.

An initial reply says that eliminativism can be safely ignored as an idle proposal, as it is not *psychologically* possible for us to control our feelings in the way it requires. The reply, though initially plausible, in fact faces several counterexamples. First of all, there are clear cases of indirect control over our feelings that are sufficient to at least raise the prospect of eliminativism. For instance, at a societal level, we know from human history that emotional kinds of once central significance to our ways of life can fall out of

¹ Some writers on the topic, such as Meir Dan-Cohen (2008: 12-3) and Joseph Raz (2011: 233-4), would say that this view of agent-regret ignores what most interested Williams: that the relevant species of regret is self-referential in taking the agent *themselves* as its object (who they are or have become), rather than what they *did*. Even if this interpretation makes the most sense of Williams—and that is a substantive exegetical claim—it is not to the point here. My interest is in the phenomenon described, which we are calling ‘agent-regret,’ whether or not that is what Williams did or meant to mean by that term.

² This is consonant with the relationship between value and emotion in general suggested by Sigrún Svavarsdóttir (2014: 89-90) and perhaps explains John Gardner’s (2018a: 141) suggestion that “The rational case for experiencing agent-regret is that [it] befits the situation one finds oneself in.”

currency (think: ‘religious’ emotions; certain forms of shame), due to contingent shifts in how we view the world and our place within it. Or again, at the individual level, it is generally accepted that we can learn to feel or to avoid feeling certain ways in response to certain situations, at least given enough time, energy, support, and so on.³

What is more, there is a familiar phenomenon that indicates at least some feelings are directly under our control. In his celebrated 1962 paper ‘Freedom and Resentment,’ P. F. Strawson describes what he calls a “resource” that we have at our disposal when we wish to avoid the strains of involvement with a troublesome person.⁴ We use it when we suspend the *participant* (aka. *reactive*) attitudes that are constitutive of ordinary human relations and instead adopt an *objective* attitude towards them, treating them and their behaviour as something to be managed, and not taken too personally. It is hard to deny that we do indeed possess this resource, and with it, some measure of direct control over our participant attitudes. As such, it seems perfectly psychologically possible for us to generally adopt an objective attitude towards our own behaviour, thereby eschewing agent-regret for the bad things we did.⁵ This is not to say that such a state is so easily achieved, only that it is in principle possible. Presumably, this is why there is a market for pop-psychology books with titles like *No Regrets* (2022) by gurus like Dr. Bill Howatt, whose occult powers are valued in the world of HR and leadership consultancy.⁶

Famously, for Strawson, the correct reply to the eliminativist is that universalising the objective attitude is not so much psychologically impossible as *conceptually* so, in a certain sense at least. Though it is “not absolutely inconceivable,” he says, he is “strongly inclined to think that it is, for us as we are, practically inconceivable” (2008: 12). What exactly Strawson means by these remarks is a complex issue, and we do well, I think, to

³ See Rüdiger Bittner (1992: 263-4) for sensitive discussion of indirect forms of control.

⁴ For discussion of the ‘resource’ and our ability to use it at will, see Strawson (2008: 9-11).

⁵ It is probably true that agent-regret, being insensitive to fault—and so perhaps “ill will or indifference or lack of concern” (2008: 15)—is not a paradigmatic participant attitude in Strawson’s sense. Nonetheless, it is plausibly: (a) a ‘participant’ attitude in a way that mere regret is not, and more importantly, (b) a conceptual condition of paradigmatic participant attitudes like resentment, i.e. in feeling ‘resentment’ we necessarily have a normative expectation that the transgressor feels agent-regret for their manifestation of ill will towards us. Moreover, as far as I can tell, the dialectic presented here between neo-Strawsonians and eliminativists applies to agent-regret without loss. For our purposes, then, it is safe to continue to regard agent-regret as a participant attitude, albeit not one that represents the paradigm case.

⁶ For more information, visit www.billhowatt.com (accessed: March 18th 2025).

draw on Pamela Hieronymi, who in a recent and authoritative monograph on ‘Freedom and Resentment’ defends the following reading.⁷ What Strawson is saying, Hieronymi argues, is that the participant attitudes play a constitutive role in any possible framework of those relationships and attachments in virtue of which our lives exhibit the kinds of organisation that make them properly human. Thus, whilst the world under a universal objectivity of attitude may be literally conceivable, it is practically inconceivable for us, in that we can only enter it by impoverishing our lives *qua* human.⁸ This perhaps explains why neo-Strawsonian reflections on such a world often have a moralised tone, as when Susan Wolf asserts that it would be one “so cold and dreary that any but the most cynical must shudder at the idea of it” (1981: 391).⁹

Against this version of Strawson, there has emerged a broad coalition of those who claim not only that we *can* (re)imagine our attachments as recognisably human without the participant attitudes, but indeed that we *should* do so. This includes both moral responsibility skeptics like Derk Pereboom (2001: 199-213), Tamler Sommers (2012: 174-88) and Per-Erik Milam (2016: 107-13), and non-skeptics like Gary Watson (1987: 284-6) and David Goldman (2014: 7-10) alike. The basic vision is a progressive one: since attachments so (re)imagined would be essentially free of any reactive or punitive tendencies, they may express, as it were, a humanity even more humane than our own. Something like this is voiced by Martin Luther King Jr. in his 1964 Nobel Peace Prize acceptance speech: “[M]an must evolve for all human conflict a method which rejects revenge, aggression and retaliation.”¹⁰ Seen from this angle, the neo-Strawsonian

⁷ Hieronymi’s treatment of the relevant remarks can be found at (2020: 24-9). For more on the wider metaphilosophical ‘social naturalism’ expressed by these remarks, see (2020: 54-62).

⁸ The neo-Strawsonian view receives perhaps its most sophisticated development in the work of R. Jay Wallace. He says (2014: 30) that “the emotions of retrospective assessment need to be seen as part of a larger syndrome of attitudes and emotional tendencies, which together count as cases of valuing or (more specifically) attachment.” Of attachment, he says (2014: 31) that “It is the main source of meaning in human existence, and a constitutive part of human flourishing or well-being.” In this way, Wallace (*ibid.*) claims to find “the deeper rationale for the emotions of retrospective assessment,” namely, that “our lives would be immeasurably impoverished if we were no longer susceptible to them.”

⁹ Strawson is, for the most part, careful not to use moralised language, though he does say that such a world would entail “human isolation” (2008: 12).

¹⁰ Available at <https://www.nobelprize.org/prizes/peace/1964/king/facts/> (accessed: March 18th 2025).

insistence on the practical inconceivability of universalising the objective attitude appears as a basically conservative inability to think beyond actually existing social practice. Given the imaginability of alternatives, however utopian, it will not do to baldly assert that attachments must involve participant attitudes if they are to count as properly human—a move that gets the inconceivability claim, in effect, only by fiat. It thus seems that there is no serious conceptual barrier to eliminativism about agent-regret either.

On closer inspection, we see that neither the psychological nor the conceptual reply to the eliminativist is especially promising. This presents a deep challenge for those, like me, who perceive that there would nonetheless be something missing, something awry, with a world where it is not in general fitting to feel agent-regret over the bad things we did. In this chapter, I outline a novel argument for this position. In contrast to dominant strands of anti-eliminativism in the literature, however, my argument develops out of a basic sympathy with the eliminativist objections just presented. Much of the novelty of the argument lies in what it suggests the problem could be with ridding ourselves of agent-regret, if not a psychological or conceptual one.

Here is a roadmap. In section 3.2, I state the core argument and proceed to defend its first premiss, namely, that the adequate explanation of reasons for action presupposes that we are engaged with the value of our own lives as actually led. I move on in section 3.3 to defend the second premiss, which says that for valuers such as we are—limited by our human bodies—to engage with the value of our own lives as actually led we have to be emotionally vulnerable to that value. Section 3.4 clarifies why this emotional vulnerability should, in the negatively valenced case, be identified with agent-regret. This allows me, in section 3.5, to say what I take to be the problem with eliminating agent-regret: that doing so would leave us without any explanation as to why we have reasons for action at all. The result, I suggest, is that the practising eliminativist would be subject to a type of inconsistency that is neither psychological, nor conceptual, but *Moorean* in character.

3.2. Explaining Reasons for Action

Our responsiveness to reasons is a basic characteristic of our form of life. This makes it important to us that the behaviour of ourselves and the people around us is guided, or

at least represents an attempt at being guided, by whatever normative reasons prevail in the circumstances at hand. Our lives would appear quite alien to us were no fact ever to count as a normative reason for or against anyone's doing anything.¹¹ For a fact to count as a genuine reason for someone to act, there needs to be an adequate explanation as to why it speaks in favour of that person's acting in that way. Absent such an explanation, the fact may exist, though the reason does not.

Simply put, the core argument of the chapter is this. For our reasons for action to have an adequate explanation, we must have a general tendency to feel agent-regret for failing to satisfy our reasons, making it apt to feel in that event. We therefore cannot do away with agent-regret, lest facts no longer count for us as reasons to act. Strictly speaking, of course, this makes the force of the argument hypothetical, leaving it open to ask why we should care whether we ever have any reasons for action at all. The answer I prefer, as just indicated, is that our lives would become quite unrecognisable to us if we did not. Whether it is in any way sustainable to claim that this is immaterial is, though interesting, beside the point here. What matters for our purposes is that eliminativists about agent-regret do not believe it is immaterial, since they themselves, as we shall see, appeal to reasons for action in their revisionary practices.

We can state the core argument more precisely, as follows.

P1 For there to be an adequate explanation of our reasons for action, we must be able to engage with the value of our lives as actually led.

P2 To be able to engage with anything that matters, we must be emotionally vulnerable to those things.

C For there to be an adequate explanation of our reasons for action, we must be emotionally vulnerable to our lives as actually led.

In this section I aim to defend *P1*, deferring defence of *P2* to the next section. The defence of *P1* proceeds by inquiry into how we explain why facts count as reasons for us to act, and what has to be the case for these explanations to be adequate. To begin to see how this could depend on our engagement with our lives as actually led, as *P1* claims,

¹¹ Hereafter, for brevity's sake, I often speak of 'reasons for action' or even just 'reasons' unless otherwise specified. I understand 'action' in a broad sense to include omissions.

let us think about explanation of reasons as they arise in ordinary conversational contexts.

ASH: 'Why are you making sushi?'

BEN: 'Because my friend is coming over.'

The context here should be clear enough. Ben is doing something. Ash asks Ben for the reason why he is doing it. Ben responds by stating a fact, which he holds out as his reason. Let us now imagine that Ash is unsatisfied by Ben's response. She does not doubt that Ben states a fact, but she doubts whether that fact genuinely counts as a reason for him to do as he does. So she asks Ben for an explanation. Ben responds in a familiar way.

ASH: 'Sure. But why is that a reason for you to make sushi?'

BEN: 'Well, our friendship is important. I guess sometimes that means we do nice things for each other, like making sushi.'

In responding to Ash, Ben refers to his friendship and its importance. This is typical: when we try to explain why some fact counts as a reason for someone to act, an appeal to the things that matter is never too far behind. That is because an adequate explanation establishes some connection between the action and the things that matter, such that we understand how the relevant fact favours the person's doing the action. When I speak of 'things' that 'matter,' I mean whatever admits of a response that is appropriate to it, that is, whatever is capable of being responded to for its own sake: activities, objects, people, practices, relationships, projects, ideals, and so on. The connection to be established is that the action either (a) is itself an appropriate response to what matters, or (b) makes some contribution—instrumentally or constitutively—to some action that does. The fragment above is a case of (a). Ben is saying that, in making sushi for his friend, he is responding to an important friendship for its own sake. An example of (b) would be the following.

ASH: 'Why are you going to the supermarket?'

BEN: 'Because my friend is coming over.'

ASH: 'Sure. But why is that a reason for you to go to the supermarket?'

BEN: 'Well, by going to the supermarket I can get the things I need to make sushi for my friend. And as I said, our friendship is important, which sometimes means we do nice things for each other, like making sushi.'

Here, Ash wants an explanation as to why the fact that Ben has a friend coming over counts as a reason for him to go to the supermarket. Ben responds that doing so would contribute instrumentally to another action—making sushi for his friend—which in turn amounts responding to his friendship for its own sake. This is how Ben establishes the connection between his action and the things that matter, in this case, his friendship.

So far, the explanation of a reason in the end appeals to how the action it favours responds non-instrumentally to what matters. That alone, however, does not obviously secure the adequacy of the explanation. Plausibly, for that, it needs to be not only that the action is or contributes to a non-instrumental response to what matters, but that such a non-instrumental response, if done well, would itself be of value.¹² And unless the value of our non-instrumental responses can be explained in turn, the explanation of our reasons in its terms seems inadequate. It is important here to be clear about the value in need of explanation. Presumably, the value there would be in our non-instrumental responses if done well is value that would be lost if done poorly. So, what needs to be explained is the value of responding non-instrumentally to what matters such as would be lost if done poorly.

Explanations for the value of our responding non-instrumentally to what matters typically appeal to the value of what matters. One traditional view takes what matters to be of value in and of itself, or in G. E. Moore's phrase: 'good *simpliciter*.' In Moore's understanding, if what matters is good *simpliciter*, then it is so completely independently of any potential for being appreciated or appropriately responded to by those capable of doing so.¹³ The value of Holbein's paintings would remain, on this view, even if humans went extinct overnight, leaving Holbein's paintings intact. I must admit that I struggle

¹² The view that what we have reasons for action if, only if, and because, it would be good if we did those actions is widely held, and has recently been labelled the "value-based theory of practical reasons" by Benjamin Kiesewetter (2022: 27).

¹³ See Moore (1922: §50).

to properly grasp any notion of ‘value’ such that it could exist independently of its being appreciable by valuers, so I put the classical Moorean view to one side. On a subtly but importantly different view, the value *simpliciter* of what matters cannot exist apart from being appreciable by valuers, but does not depend on the value of those valuers, their appreciation, or anything else. Instead, our appreciating what matters is good because in doing so we partake of, participate in, or otherwise instantiate the value *simpliciter* of what matters. Thus, on this view, our responding non-instrumentally to what matters is explained as itself of value *simpliciter*.¹⁴ This perhaps improves on the Moorean view, but the issue, as I see it, is that it makes what it is for our non-instrumental responses to be valuable nothing more than their possession of the property of being good *simpliciter*.¹⁵

I do not see why we should accept any view with this implication. I prefer to say that what it is for our non-instrumental responses to be valuable is that they embody successful *respect* for or *engagement* with what matters.¹⁶ We show respect when we act so as not to destroy (and perhaps to preserve or protect) what matters, which is, as such, the minimal way of responding to it non-instrumentally. For instance, at the National Portrait Gallery, we might respond to a Holbein by being careful around it, imploring others not to touch the canvas or vandalise it—all for the sake of the painting itself. More positively, what matters call for engagement with it for its own sake, a type of response whose depth and complexity may be thought to separate us from other forms of life. The mark of engagement is the reflexive interplay of feelings, thoughts, imagery and action in which our critical faculties are most fully manifest. To return to the gallery, we may respond to the Holbein by considering the subtlety of technique, the symbolism, the politics of the Tudor period, and—letting all of that significance play within us—being moved by it anew. Such engagement may go better or worse, but either way, our lives are led in so doing, accruing the specific content they do. In protecting the things that matter, respect safeguards the general possibility of leading life, from which it derives its point. Plausibly, then, the value of responding non-instrumentally to what matters lies first and foremost in how doing so is what our lives as actually led consist in.

If this is correct, it is hard to see how the goodness *simpliciter* of what matters could explain it. What matters figures not so much as instantiating some property, goodness

¹⁴ See Donald H. Regan (2004: 221) for an example of such a view.

¹⁵ See L. Nandi Theunissen (2025: 234-5) for this pattern of argument.

¹⁶ For fuller discussion of respect and engagement, see Chapter Four.

simpliciter, that our non-instrumental responses merely partake of or not. Rather it figures as a potential to be realised or lost through our non-instrumental responses to it, which, in being able to go better or worse, can be meaningful for us, constituting the specific content of our lives as actually led. As such, it seems able to explain the value of non-instrumental responses if done well, such that, if done poorly, the same value is lost. I find it plausible, therefore, that what matters is not good *simpliciter*, but good *for* us, of *benefit* to us, in that responding to it non-instrumentally, however well or poorly, is what our lives as led actually consist in. This is what I believe explains the value of responding non-instrumentally, value that is lost if done poorly. What explains it is not the value *simpliciter* of what matters, but its value for us, its benefit to us.

Of course, as far as the explanation of reasons goes, this only pushes the problem one step back. We now need to know: what explains the value of what matters, in the sense? For unless this can be explained in turn, the explanation of the value of non-instrumental responses falls short, and with it the explanation of reasons. Another traditional view appeals at this stage to the value of valuers themselves. What matters is good for us in that engaging with it non-instrumentally is what constitutes our lives as led, and what explains that value, on this view, is the value of we who lead those lives. Perhaps the most familiar version of this view takes the value of valuers to be value *simpliciter*, or in the Kantian idiom: ‘unconditioned’ or ‘absolute.’¹⁷ I am not sure it can be made any more convincing to appeal to the value *simpliciter* of valuers than that of what matters, not least because valuers themselves matter. Nonetheless such views have remained popular, perhaps in part due to the palpable threat of vicious regress. As I have mentioned, what is valuable *simpliciter* does not depend for its value on that of anything else, it stands in need of no further explanation. However, worries about vicious regress can be assuaged without appeal to value *simpliciter*, since, as L. Nandi Theunissen (2018: 368) has argued, “A chain of dependence between relational values comes to an end with a reflexive relation [...]” Thus, it is possible to appeal to the value that valuers have *for* something else, rather than in and of themselves.

In my view, this way of proceeding is not so intuitive. We need not invoke the value of valuers, whether value *simpliciter* or otherwise, to explain the value of what matters.

¹⁷ Versions of this view are found in Christine Korsgaard (1983: 177-84; 1986: 190-7), Joseph Raz (1999: 296-7; 2001: 145-51) and J. David Velleman (1999: 609-11; 2008: 210-11).

In this respect, I agree with more recent remarks from Theunissen (2023: 232), who says “I no longer think that a datum about our value falls out of the structure of evaluative explanation.”¹⁸ We are trying to explain why what matters is good for us in that engaging with it non-instrumentally is what our lives as actually led consist in. The obvious answer, surely, does not appeal to our own value, but the value of our lives as actually led. One possible, if not popular, way of proceeding is to claim that this value is value *simpliciter*. I put it to one side, however, because it is no more plausible than the appeal to the value *simpliciter* of what matters, not least because our lives as actually led themselves seem to matter.¹⁹ More plausibly, our lives as actually led are good in the sense of being good for us. Of course, this can easily make it appear that the value in question is the value of the good life, of our own wellbeing. But for present purposes, this will not do. What matters is good for us in that responding to it non-instrumentally, *however well or poorly*, is what our lives as led consist in. So, if *this* benefit is to be explained by the value of our lives as led, then that cannot be identified with the good life, since we might actually be leading our lives very poorly, i.e., our lives might be going quite badly. The relevant value of our lives as led is value they must have even if we are not actually leading them very well.

I believe that the relevant value is the value our lives as actually led possess in virtue of being amongst the things that matter. As with anything that matters, they admit of being responded to non-instrumentally. In leading them—now better, now worse—our lives as actually led come to possess their own distinctive quality. They have their ups and downs, a narrative shape, a place in history, and so on, that we can interpret, reinterpret and respond to in the complex and multimodal way that amounts to involving ourselves in them for their own sake. Like anything that matters, our lives as actually led are good in the sense of good for us, and that good, that benefit, lies in how engaging with it non-instrumentally is what our lives as actually led consist in. It is the value of our lives as actually led in this sense that explains the value of what matters more broadly.

¹⁸ In her earlier work, Theunissen (2018; 2020) was amongst those who sought to explain why what matters is good for us by appeal to our being good for something else, namely, our own lives. She later dispensed with this aspect of her view in light of comments from Kenneth Walden (2021).

¹⁹ For further reasons against such a view, see Theunissen (2018: 369).

Have we not, once more, simply pushed the problem one step back? It may seem that we now need to explain the value of our lives as actually led. For unless this can be explained in turn, it may seem that the explanation of the value of what matters falls short, and with it the explanation of the value of responding non-instrumentally to what matters, and ultimately the explanation of reasons. The worry overlooks the fact that the explanation has already been given. In general, we said, the value of what matters is explained by the value of our lives as actually led in engaging with what matters for its own sake. Our lives as actually led matter, and so their value, too, is explained by the value of our lives as actually led in engaging non-instrumentally with those lives. This renders non-vicious the regress in the structure of evaluative explanation with the sort of “reflexive relation” described above by Theunissen (2018: 368). Perhaps it will seem that the proposed relation, though reflexive, cannot be explanatory, because the relata are identical. But they are not: in engaging non-instrumentally with our lives as actually led, we actually lead our lives, and so the life that we respond to cannot be the same as the life that includes the response. They are, as it were, substantially different texts, and correspondingly, their value for us differs.

3.3. Value, Attachment and Emotional Vulnerability

In this section, I aim to defend *P2*, the premiss that to engage with anything that matters, we must be emotionally vulnerable to it. I will try to make plausible two observations about the conditions of our engagement with worthwhile things, with a view to saying something about what we must be like as valuers if there is to be an explanation of our reasons for action. The first observation concerns the existence of a quite general threat to our ability to engage with things, and what it takes for us to defuse it.

When we go about our lives, we do so, for the most part, quite preoccupied by those things—the activities, relationships, projects, and so on—that really matter to us. This is to be expected. Our immersion in these things is what our lives as led consist in, lending those lives their distinctive shape and quality. But as reflective beings, we can, in thought, take a step back from what matters to us, revealing a world replete with worthy things far in excess of those we happen to be involved with. There are indefinitely many worthy things that could have ended up mattering to us instead, such that our whole lives would

have been organised around those things, rather than the ones familiar to us from the lives we have ended up leading. The things that matter seem not to admit of meaningful comparison—or not, at least, as they bear on the orientation of a whole life. In themselves, they do not make it better to pursue one rather than another. Just as there is nothing about apples and oranges that could settle once and for all which is the superior fruit, the same is true of a career in law or in medicine, the ideals of justice or charity, or intimacy with this or that person.

I claim that for valuers such as we are, a world replete with plural and incomparably worthy things poses a threat to our ability to engage. The threat will be familiar from the story of Buridan's donkey, who, being caught between a source of food in one direction and water in another, perishes due to an inability to focus on either thing over the other. Analogously, I claim, if we are not to become paralysed in a like fashion—that is, if we are ever to engage with anything worthwhile at all—we must ourselves be able to focus on some things over others, in a manner that accords with how they matter. Or, at least, we must have some such capacity *if* there is to be any explanation as to why facts count as reasons for us to act. We saw why in section 3.2. To explain why facts count as reasons, we immediately appeal to the value of responding non-instrumentally to what matters, which, of course, cannot exist if there can exist no such responses—engagement being principal amongst them.

One such capacity will be familiar enough. Amongst the myriad things that matter, some of them, it seems, get taken up into our lives. When this happens, they become the objects of what I will call *attachment*—though some prefer to say we come to 'care' about, have a 'commitment' to, or simply 'value' those things.²⁰ In any case, I am interested not so much in the label as the phenomenon. Once we are attached to whatever we are attached to, they take precedence over the things we are not attached to, however much they may in fact matter, and however cognisant of that fact we may be. As I write this sentence, for example, I am aware of many other worthy things I could be doing instead: composing music, baking a cake, canvassing for my party, and so forth. And yet, I continue to write this paper, and it is probably fine for me to do so, because philosophy is part of my life in a way that music, baking or party politics are simply not. It is plausible that our attachments are uniquely poised to allow us to focus on some worthy

²⁰ See Chapter Four for more detailed discussion of attachment.

things over others, and so, I argue, we must have a capacity to form them, at least if there is to be any explanation of our reasons for action.²¹

I note that the plural and incomparable nature of what matters might not have been a threat were we not, as valuers, such as we in fact are. Consider that if Buridan's donkey had been big enough to span the distance between the sources of food and water, then it would have been able to consume both at once instead of being paralysed between them. Similarly, I believe, if we were big enough, there would be no threat (or no general threat) of being paralysed between engaging over here with one worthy thing and over there with another. We would thus have no general need for the capacity of attachment. According to Byong-Chul Han, this is the point of the unthinkable dimensions of the wondrous animals who inhabit the anecdotes of Zhuangzi, the 4th century BC Daoist philosopher.²² Zhuangzi teaches that those who are not attached to particular things in the world are "beyond the possibility of loss" (2023: 15). But so long as we are "*smaller* than the world," and so amongst worldly things, we "will be affected by *care*" (2023: 16). In this sense, "Someone who is as *big as the world* will not be hindered or impeded by anything *in* the world" (2023: 62). Thus, for Zhuangzi, says Han, the ethical significance of being big is "to de-differentiate oneself into an *impartial friendliness*" (ibid.). If we were as big as the world, that is, we would have no need of the partiality of attachment, in which case it could well be better for us to dispense with it. Though it perhaps strains the imagination, it is a thought I can both understand and appreciate. Perhaps it is even true. But even if it is true, it has no practical purchase on us, for the simple reason that we are not in fact as big as the world. Hence, it cannot be that for us *such as we are*—with our human bodily dimensions—we could or should heed the ethical injunction to stop caring about particular worldly things.

²¹ The relevant capacity to form attachments need not involve anything like choosing them or setting them for ourselves in any way, since that is not needed to defuse the threat. This chimes with Raz's description of the capacity to form 'goals,' his preferred label for the phenomenon. He says (1986: 290-1), "Some of these goals a person may have adopted deliberately, some he may have chosen. Others he may have drifted into, grown up with, never realized that anyone can fail to have them, etc. It makes no difference from our point of view which is which."

²² Such as the fish named Kun (鯀) who transforms into the bird named Peng (鵬), both of whose dimensions are, in Brook Ziporyn's interpretation, on the order of "who knows how many thousands of miles" (2020: 4).

Let us briefly recap the development of section 3.3 so far. I have been trying to make plausible the observation that for valuers such as we are—who are confined to a human body with its limited spatial extension—the plurality and incomparability of the things that matter poses a general threat to our engagement with those things, a threat that our capacity to form attachments is uniquely poised to defuse. With the help of the section 3.2 suggestion that our engagement with what matters is amongst the conditions of the explanation of reasons, I used the observation to draw an intermediary conclusion: the explanation of reasons depends, *inter alia*, on our capacity to form attachments.

I now offer a second observation, which, together with the intermediary conclusion, will enable me to vindicate P2. The second observation is that it is hard to imagine how we could form attachments at all without our being in some way selectively vulnerable to the objects of our attachments. Without something like this, I claim, it will remain a mystery how we gravitate towards some worthy things and not others, so that we can become attached to them in the first place. We must have some such vulnerability if we are to become attached to and engage with the things that matter, as we must, as I have argued, if there is to be any explanation for our reasons for action.

One such vulnerability will be familiar enough from our own case. We seem to be differentially *emotionally vulnerable* to different things that matter. I think that what this means in psychological terms can be left fairly intuitive for present purposes. It does not need filling out much more than that our cognitive systems are so constituted as to be able to recognise, however inchoately, that certain things matter in certain ways, and that our motivational systems are so constituted that we are liable to be drawn towards some of the things we recognise as mattering and not others. When certain things draw us in, we react with a positive hedonic valence when events benefit them—pleasurable feelings of attraction, satisfaction, and so on—and a negative valence when events disbenefit them—painful feelings of aversion, discontent, and the like. It also involves wanting, in a broad sense of ‘want,’ to engage with those things in various ways, which can also be to their benefit or detriment. The pattern of our emotional vulnerability to the things that matter is under our control only very obliquely if it is at all. Which things draw us in, in what way, and how much, will differ depending on who we are: our tastes, background, personality, aptitudes, and so on. It is plausible, I think, that our emotional vulnerability in this intuitive sense (or something like it) is uniquely poised to deliver us

into our attachments. Given that our capacity to form attachments is uniquely poised to defuse the threat to engagement, it follows that *P2* is true.

Together with the last section, this vindicates the core argument of the paper. It says: for valuers such as we are, with our finite human bodies, the explanation of our reasons for action depends on emotional vulnerability to the value of our own lives as actually led. Or, in more precise terms, once again:

P1 For there to be an adequate explanation of our reasons for action, we must be able to engage with the value of our lives as actually led.

P2 To be able to engage with anything that matters, we must be emotionally vulnerable to those things.

C For there to be an adequate explanation of our reasons for action, we must be emotionally vulnerable to our lives as actually led.

This argument brings us well within the vicinity of the eliminativist challenge I set out to address. Let me begin to join some of the dots. We explain our reasons for action by pointing to how those actions either are or contribute to valuable non-instrumental responses to what matters, a value such as is lost if those responses fall short. Such value depends on the value of what matters, which depends on how our responding to it non-instrumentally (however well or poorly) constitutes our lives as led, the value of which depends in turn on how our engaging with them further constitutes our valuable lives as led. But to engage with our valuable lives as actually led, however well or poorly they are led—that is, whether our responding non-instrumentally to what matters realises or loses value—we have to be emotionally vulnerable to them. This emotional vulnerability in general involves reacting with a positive hedonic valence to what benefits its objects and a negative valence to what disbenefits them. What benefits or disbenefits our lives as actually led, what constitutes their ups and downs, is our non-instrumental responses to what matters, insofar as they realise value or lose it. To explain our reasons for action, then, we must react pleasurably to actions of ours that (contribute to) realise value, and painfully to those that (contribute to) its loss. Thus, if our reasons for action are to have an adequate explanation, it must be, *inter alia*, that we feel painfully if we fail to satisfy them.

I will now close the section with some objections. To start off with, we may worry that neither the capacity to form attachments nor our emotional vulnerability to what matters is functionally unique in the way the argument requires. That is: if there were some other way than our attachments for us to focus on some worthy things over others, or if, even if not, there were some other way than our emotional vulnerability to become attached to the worthy things we do, then these alternative capacities would suffice to secure the explanation of reasons. The obvious candidate here is some sort of random value generating process, e.g. a coin toss or a dice roll. The problem with such processes, however, is that although they may well help us attend to certain things over others, they do not do so in accordance with or guided by the way they matter. They reflect, fundamentally, a distance or dissociation from the things that matter that can hardly serve as a condition of attachment or engagement with them for their own sake.²³ On the question of whether anything else about us serves, or could serve, the relevant functions, I regard that an empirical matter to be settled by the life or social sciences, or indeed works of speculative fiction. I do not want to suggest that developments in these domains might not uncover concrete alternatives. But until such time, it seems plausible to maintain that our attachments and emotional vulnerability are functionally unique in the way the argument requires.²⁴

Another objection is as follows. It may be thought implausible that our normative reasons for action could depend on emotional vulnerability to our lives as led, in that this would, if true, make our reasons depend on the subjective motivational sets of individuals in a manner at odds with robust conceptions of normativity. We can bring

²³ Something like this is the optimistic reading of George Cockcroft's frequently disturbing 1971 novel *The Dice Man*, in which the protagonist, a depressed psychiatrist, begins to let his life and actions be determined by the roll of a dice, with destructive consequences. The pessimistic reading is that the protagonist does so intelligibly, because under the distinctive conditions of late modernity, this is the only way for his life and actions to have meaning any more. The author's toying in earnest with the availability of the latter reading goes some way, I think, towards explaining the somewhat dated feel of this novel today.

²⁴ Recent empirical psychology can be marshalled to support the functional uniqueness claim. For example, Heidi Grant and Laura Gelety (2009) recognise so-called 'life goals,' a phenomenon much like attachment. Emotional vulnerability can be borne out, in turn, by understanding 'life goals' as a species of 'goals' on the hierarchical model of approach-avoidance motivation, on which the adoption of goals presupposes our faculty for affective responses to various stimuli. See Andrew J. Elliot and Daniela Niesta (2009) for the latter model.

this point out with reference to depression. Depression in real life is of course a complex, personal and phenomenologically varied clinical condition, which I neither intend nor am qualified to talk about here. But we can, as philosophers, distinguish the concepts of two caricatures of depression, based on how being in that state relates to the rationality of suicide. In type-(i) depression, a depressive episode makes sense as a response to life going badly, where its depth reflects just how badly life is going. Assuming that suicide expresses a certain depth of depression, someone whose life is going abysmally—and this is easiest to imagine as inflicted upon them by a malicious party—may thereby have a reason to commit suicide, if they are type-(i) depressed. This stands in contrast to type-(ii) depression, where we become to some extent disinterested or dissociated from our lives as led, not (or not only) on the level of belief or judgment as to their mattering, but on a deeper motivational level. The result is a generalised disengagement, at the limit a complete cessation of involvement in our lives as led.²⁵ As an empirical matter, I am not sure whether it is possible for anyone to find themselves totally in the grip of type-(ii) depression. But that is beside the point here. For present purposes, what is important is that such a person could have no reason to commit suicide, for, if my argument is correct, they would have no reasons for action at all.

With this distinction in place, we can state the objection. The worry is that this seems to hold normativity hostage to the good psychological health of individuals, in that their sinking into total type-(ii) depression would destroy our having reasons at all. I believe the worry is misplaced. From the beginning, the point has been to explain why facts count as reasons for us to act a certain way, when they do. The argument says only that, for a given fact and person, the explanation of *that* person's reasons requires that *that very* person must be emotionally vulnerable to the value of *their own* lives as actually led. Thus, it only implies that if someone were to sink into a totalising type-(ii) depression, or to otherwise find themselves numb to how their lives are going, then that would deprive them of their reasons for action. It does not imply, however, that this situation would necessarily have any bearing on the existence anyone else's reasons for action. Though someone totally type-(ii) depressed may cease to have reasons to act, whether to commit suicide or anything else, those around them do not, so long as they

²⁵ See Léa Salje (2023) for clinical sources that point to real life depression being closer to the type-(ii) caricature, as well as an interesting proposal as to how it differs from the Epicurean ideal of *ataraxia*.

continue to be moved by their own lives. That includes reasons they may have to help those in the oblivion of type-(ii) depression, by helping to restore their attachment to their lives, the value of which have become so motivationally inert for them.

This may still be thought unattractive because it severs the connection, introduced in section 3.2, between the explanation of our reasons and the value of what matters. To see this, consider that on my view, it may be that the things that matter do not give some people (e.g. those suffering a total type-(ii) depression) any reason at all to respond non-instrumentally to them. Though this is indeed a troubling possibility, I believe we have the resources to make sense of it. Imagine sitting a dog before a gramophone and playing it Scriabin's *24 Preludes*. It is not that the dog cannot respond to it at all: it does, after all, receive the notes through its ears, one environmental stimulus amongst many. But what it cannot do is respond appropriately to the value of the *Preludes*. Of course, the fact that the *Preludes* do not give the dog any reason to appreciate them does not, and should not, make us fret about the power of the *Preludes* to give the rest of us reasons to appreciate them. We are inclined to say instead that the *Preludes*, though beautiful, are lost on the dog. I think the same should be said of someone totally in the grip of type-(ii) depression. The tragedy is that the value of their own lives becomes lost on them.

3.4. The Primacy of Activity in Life

Before concluding, I want to say something about the appeal I make to the value of our lives. I have placed an emphasis on our lives *as actually led*. We may wonder: what is the special significance of our lives as actually led, over and above our lives *we actually have*? After all, the lives we actually have seem to consist of more than just how we actually led them to date. In focusing narrowly on our lives as led, it may seem that I have ascribed value to those lives based on a very impoverished view of their content. Is it not the lives we actually have that is of more fundamental significance?

I agree that the lives we actually have are of significance. Their value, their quality, is plausibly a function of how they are going *in toto*, that is, insofar as we are *active*—what we do or do not do—and *passive*—what does or does not happen to us. We are the agents of our lives, we lead our lives, insofar as our activity is or contributes to our non-instrumental responses to the things that matter, however well or poorly we may do so.

Conversely, we are the patients of our lives when events transpire that are in some way beneficial or detrimental to how our lives are going, but which are not the result of any activity on our part in the relevant sense. For instance, lightning may strike someone's garden shed or they may lose their job due to sectoral collapse during a global pandemic. Or, more happily, they may win the lottery. It is plausible that such events are amongst the determinants of the lives we actually have or end up with, even as we are, in a sense, the passive recipients of them.

I deny, however, that the lives we actually have are of fundamental significance. It is the active side of life, our lives as actually led, that matters more fundamentally. To see this, consider, first of all, that whether what merely happens to someone is a boon or a bust depends on its interacting, and on how it interacts, with the wider context of their plans. New information about what a person was up to at the time of the event can easily make lightning strikes, job losses and lottery wins take on a new complexion. Perhaps, for example, they had plans to remove the garden shed and the lightning strike ends up saving them time, money and energy on a demolition contractor. Perhaps they had been stuck in a dead-end job but could not muster the resolve to find a new job, for weakness of will. Or perhaps winning the lottery makes it harder to kick their gambling addiction, as they had been trying to do for many years. The fact that the meaning of what merely happens to us exhibits this indeterminacy absent reference to the wider context of what we were up to suggests a dependence of the passive on the active in life.

It may be objected that events that happen to us that impact our brute biological needs can influence how our lives are going quite independently of the wider context of our activity. An uncontroversial list might include, for example, events that affect our state of nourishment, shelter, physical or psychological health, and so on. To this line of thought, Raz (1986: 306-7) offers the following thought experiment:

think of a person who is entirely passive, and is continuously fed, cleaned, and pumped full with hash, so that he is perpetually content, and wants nothing but to stay in the same condition. It's a familiar imaginary horror. How do we rank the success of such a life? It is not the worst life one can have. It is simply not a life at all. It lacks activity, it lacks goals.

The point is that however much merely happens to a person that looks *prima facie* beneficial or detrimental to their life, if they have no active engagements to give context to these events—if they have no life as led—then they have, strictly speaking, no life to benefit or suffer detriment. Any existence of a person such that things merely happen to them but without any activity on their part cannot be understood as a life. We actually have no life except insofar as we have a life as actually led. Activity, unlike passivity, is in this sense not just any old constituent of our lives, but an *essential* one. The active takes primacy over the passive in life in this sense. Hence, as Raz says, “the concentration on activity is meant to flow from the very notion of a life” (1996: 3). The relevant notion is not (or not just) that of organic or biological life, though being alive in this sense is, presumably, a material condition or basis of it. Rather, it is the one that “forms our self-consciousness” and “from which we cannot escape” (1996: 3n). What Raz has in mind, I take it, is the sort of life that we are each of us intimately acquainted with, one that has a biography, a narrative arc, a history of successes and failures. And that is the sort of life in which the active is an essential ingredient, the sort of life we cannot actually have except insofar as we actually lead it.

This explains my emphasis on our lives as actually led in explaining our reasons for action. Recall from section 3.2 that, on my view, what it is for our activity to be of value is that it is what our lives consist in. It is this value of our non-instrumental responses that the value of what matters is in turn said to explain. But this value cannot be value that our responding non-instrumentally to what matters just happens to have, that is, just when our lives happen to consist in those responses. Otherwise, even if what matters is genuinely valuable, our responding non-instrumentally to it might not be, because our lives might happen not to consist in those responses. And the value of responding non-instrumentally to the valuable things that matter is not contingent in this way. Rather, it is value our responding non-instrumentally necessarily has as an essential constituent of our lives, lives we would not have but for our responses having such value. The value of our non-instrumental responses is explained by the value of our lives insofar as they are essentially constituted by such responses, our lives in their active dimension. Hence my appeal to the value of our lives as actually led, not just the lives we actually have.

3.5. Concluding Remarks

The argument of this chapter, if correct, licenses the conclusion that if there is to be any explanation as to why we have reasons for action, then it must be, *inter alia*, that we feel painfully if we fail to satisfy them. This bears directly on the eliminativism about agent-regret I set out to address. Recall that the eliminativist doubts that is ever really fitting to feel badly about the 'bad' things we did, the occasion of a loss of value we occasioned, enjoining us to stop with the irrational self-flagellation by doing away with agent-regret altogether. The standard reply, I said, is to suggest that actually adopting eliminativist practices would be inconsistent, either on psychological or neo-Strawsonian conceptual grounds.

As we saw, however, the charge of inconsistency on either of these grounds does not look possible to maintain. For it seems, first of all, to be within our psychological powers to generally adopt an objective attitude towards our own behaviour, and moreover, that to do so does not obviously make it incoherent to say that we are living a properly human life, with properly human attachments. Thus, with nothing for us to lose but the needless pain of agent-regret, this is just what the eliminativist would have us do. This need not, of course, imply that we should cease to properly appreciate the badness of what we did or its normative import going forwards: the reasons that fact that we occasioned a loss of value gives us to learn from it, to change, and in some cases, to make amends. For if it did, the eliminativist would be hard pressed to maintain the claim to a properly human life. It only recommends that we properly appreciate the bad things we did without feeling badly about them, that is, by grasping them in cold, emotionless evaluative judgment. An adequate reply to the eliminativist therefore requires us to point to some inconsistency about adopting their practice that is neither psychological nor conceptual in Strawson's sense.

On the view proposed here, the eliminativist about agent-regret is subject to a type of *Moorean* inconsistency. To see this, imagine a practising eliminativist in the moments after they have done something that is in fact bad. As we just saw, they grasp this fact in emotionless evaluative judgment, take it to have normative import, and yet do away with feelings of agent-regret. But though it may seem innocuous at first, this combination of attitudes is weird in much the same way as it would be weird to judge a proposition true whilst at the same time not believing it. In the latter case, it is weird because no one can

ascribe the normative property of truth to a proposition without implying that it is an apt object of belief: a reason of fittingness that, in not believing it, they weirdly flout. It is equally weird, I claim, for the eliminativist to judge that what they did was bad without feeling painfully about it. I argued that for the value there would be in acting to explain our reasons to do so (i.e. for it to be a normative property), we must feel painfully if we fail to realise it in so acting, making us apt to feel that way. So, upon such failure, anyone who continues to take the value there would have been to give them reasons for action presupposes that it is apt to feel painfully about their failure: a reason of fittingness that, in not feeling that way, they weirdly flout. And this is exactly the attitudinal situation of the practising eliminativist. That is because, for all their lack of agent-regret, they continue to take the value lost in what they did, the badness of what they did, to support reasons for them to act: to learn from their mistakes, to change, to repair the situation, and other such distinctively human things.

My argument aims to meet the eliminativist at the same deep level on which their doubts operate. But it does so without trying to brute force their recommendations out of existence, by alleging that they are psychologically or conceptually impossible. In this respect, my reply is more modest than the other replies we considered. The force of the argument is not that the eliminativist's proposals are impossible, but rather, that they are not actionable on the basis of reasons. For even if valuers such as we are *could* adopt an eliminativist way of life and it really would be as *good* as its proponents imagine, that could not give us *reasons* to enter such a state, because doing as those reasons say to do would undermine the possibility of explaining why we have them.

Perhaps in taking the eliminativist so seriously, I already fail, like many others, to do what Wittgenstein says is so difficult to do: “[T]o begin at the beginning. And not to try to go further back” (1969: §471). For those, like Wittgenstein, of a broadly ‘naturalist’ bent in Strawson’s sense, we are, of course, to begin with the inescapable facts of human nature, including our social nature. Thus, in letting these facts get called into question, I am, in the naturalist’s eyes, engaged in the futile attempt to begin further back than the beginning.²⁶ On the contrary, it strikes me that the beginning is located wherever we can first sensibly ask for reasons. And we can sensibly ask for reasons where we can imagine

²⁶ Wallace (2014: 32) puts the point nicely when he says: “there can be no real question for us of ceasing altogether to care about things, and the emotions to which we are susceptible in virtue of this valuing stance are in that way beyond justification.”

alternative ways of doing things. Our attachments and the emotional vulnerabilities that make them up appear to be, in this respect, rather more malleable than certain others of the naturalists' favourite practices, such as our epistemic reliance on the existence of body or on the principles of inductive reasoning.²⁷ To try to deny this by appeal to some special insight into the inescapable substance of human nature flies in the face of imagination, and reflects, it seems to me, a conservative cast of mind.

Does my view not also appeal to human nature? Well yes, it does, but only to a much thinner notion, and in a fundamentally different way. On my view, we must feel agent-regret because without it we would not be able to form the attachments we need if we are to engage with what matters, given we are such as we are: constrained to live in human bodies of finite spatial extension. And this is required, not because we would otherwise be bad or humanly deficient, as the neo-Strawsonians like to insist,²⁸ but because there could otherwise be no explanation as to why we have reasons at all. As such, my critique of the eliminativist is less that their highly progressive proposals would impoverish us than that they are overly ambitious: they seek to overcome the limits our bodies place on the pursuit of what matters. For valuers such as we are, the injunction to dispense with agent-regret is in this way *too* utopian, *too* aspirational. We cannot make ourselves more human by being more than human.

²⁷ These Humean examples are taken from Strawson (1985: 19).

²⁸ Recall for instance the moralised tenor of the anti-eliminativist assertions in Wolf (1981: 391), exposed by Sommers (2012: 166-9) as simply lacking in imagination.

CHAPTER FOUR

On the Reasons We Cannot Escape

4.1. Preliminary Remarks

In the preamble to Part Two, I outlined an account on which ordinary agent-regret acquires the desired ‘moral’ quality when it responds to failures to satisfy extraordinary reasons for action. Chapter Three defended the claim that we are apt to feel agent-regret for failing to satisfy our reasons for action *at all*. I will now defend an account of what is so extraordinary about the reasons I have in mind, namely, that they have the normative inescapability and requiring force that together characterise our *obligations*. Chapters Four and Five defend accounts of inescapability and requiring force respectively.

According to the widely held view Benjamin Kiesewetter (2022: 27) calls the *value-based theory of practical reasons*, for anything anyone can do, they have a reason to do it if, only if, and because, it would be of value if they did it. I assume this view throughout. I believe this view, coupled with some plausible thoughts about the value of our actions, allows us to advance our understanding of a sense (to be explained) in which reasons are ‘inescapable,’ when they are.

I proceed as follows. In section 4.2, I lay out the independence of the value of respect for what matters from both our attachments and our conative attitudes, and distinguish the value of respect from that of engagement. Section 4.3, introduces the phenomenon of attachment-dependent obligation, exploring potential difficulties raised by this

phenomenon, in particular surrounding the unity of the sense of ‘inescapability’ at issue. In section 4.4, I suggest an interpretation of ‘inescapability’ that resolves the difficulty about the unity of the sense of ‘inescapability’ generated by our attachment-dependent obligations. Section 4.5 briefly concludes. One major upshot of the chapter is that attachment-dependent reasons cannot be explained as ‘inescapable,’ in the relevant sense—so *a fortiori* as obligations—unless the attachments on which they depend are themselves of value.

4.2. Respect, Engagement and Attitude-Independence

Suppose I am on holiday in the Himalaya, in wild, unspoilt terrain of sheer natural beauty. It would be good if I were to move through my surroundings in such a way as to preserve the structural integrity of the terrain, to sustain the peace and quiet, not to disturb the flora and fauna, and so on. In acting to preserve, protect, or at least not to destroy the unspoilt nature, I *respect* the Himalaya.

What goes for me, goes for you, and for anyone else capable of responding appropriately to the things that matter. For each and every one of us, it would be good if we respected the Himalaya, regardless of whether we are especially bothered about natural beauty. Of course, if I am on holiday in the Himalaya, I am likely to be enthusiastic about natural beauty. But that need not be so. Perhaps the built environment, man-made structures and post-industrial decay better suit my tastes, temperament and aptitudes, whereas conversely, the wilderness leaves me somehow cold. Perhaps, that is, I have an *attachment* to urbanism in a way I simply lack with the natural world. Even so, if I am on holiday in the Himalaya, it would be good if I moved through my environment respectfully, keeping my disturbance of its unspoilt nature to a minimum. And the same is true for anyone else, that is, whether or not they are attached to the Himalaya, natural beauty, or the wider natural world. To generalise: for any person, and anything that matters, it would be good if they respected it no matter whether they are attached to it. Applying the value-based theory of practical reasons, our reasons of respect are explained, in Monika Betzler’s (2015) phrase, as *attachment-independent*.

Some things it would be good if we did, but not necessarily for each and every one of us. Suppose you are also in the Himalaya, not as a tourist, like me, but as a seasoned

alpinist. It would be good if you were to respect your surroundings in the same way it would be for me. But given your love of alpinism, there are things it would be good if you did that it simply would not be if I did. Suppose, for instance, that we both notice an impressive gully rising from the foothills of a nearby peak. Given your love of alpinism, you would be remiss not to at least take a closer look at the gully, assess its features, consider possible routes, and hopefully make an ascent attempt if the conditions permit it. It would be good if you actively *engage* with this potential in the gully. By contrast, even if I can nascently appreciate this potential—or even if I would, on witnessing your head for heights and surefootedness, gain a deeper respect for your practice—I am hardly remiss not to make an ascent attempt myself. That is because alpinism is not a part of my life in the way it is for you. I have not (or not yet) been drawn into a personal history with alpinism by my tastes, temperament and aptitudes, such that it matters *for me* as it does for you. It is not (or not yet) an object of my attachment. This is why it would be good if you made an ascent attempt, but not me.

Applying the value-based theory of practical reasons, we get the result that our reasons of engagement are *attachment-dependent*. The attachment-dependence of reasons of engagement distinguishes them from reasons of respect, which are, by contrast, attachment-independent. ‘Attachment-dependence,’ here, should not be taken to mean that the fact that we are attached to something is itself our reason to engage with it. Rather, it is the fact that our engaging with it would be of value that is our reason to do so. It is just that there can be no value in engaging with anything that matters unless we first become attached to it, which stands in contrast with the value of respecting it.

Some things it would be good to do, but only given our conative attitudes—our goals, plans, intentions, desires, and the like. Suppose, as before, you are an alpinist. Though you have not yet been to the Himalaya, you have dreams of a Himalayan expedition in the coming years. You have the option of spending your annual leave training in the Cairngorms right now, in the summer, or you can await the winter, when conditions will closely resemble those in the Himalaya. Presumably, it would be good if you were to await the winter, but only given your dreams of the Himalaya. If you had no such dreams, and instead had ambitions to improve your rock climbing, or simply preferred the Cairngorms in summertime, then other things equal your awaiting the winter would be somewhat pointless. Since the value of your awaiting the winter

depends on your having dreams of the Himalaya, your reason to await the winter is, on the value-based theory of practical reasons, an *attitude-dependent* reason.

This is the ordinary way our conative attitudes give us reasons to act, when they do. The ‘attitude-dependence’ of a reason to act should not be taken to mean that our merely having or satisfying an attitude of that type is itself the reason. Rather, the reason to act is that it would be good if we did it, given our conative attitudes. Such a dependence of the value of action on conative attitudes obtains, when it does, because those conative attitudes partly constitute the way we have come to be attached to what matters, so that satisfying those attitudes amounts to appropriate engagement, which it would be good to do. Thus, all attitude-dependent reasons are reasons of engagement. Our conative attitudes do not give us reasons if they do not partly constitute our attachment to what matters, for in satisfying them we do not engage appropriately with anything that matters.

To see this, suppose, first of all, that your Himalayan dreams are part and parcel of your particular love of alpinism, constituting your attachment’s concrete history, whereas your preference for the Cairngorms in the summertime is more incidental. In that case, though your dreams give you a reason to await the winter, your preference does not give you a reason to take leave in the summer (though facts about the Cairngorms in the summertime might). That is because the dream, but not the preference, partly constitutes the particular way you have come to be attached to alpinism, such that satisfying the dream, but not the preference, amounts to engaging appropriately with alpinism. Now suppose that, after some sobering conversations, you have come to believe that you have become too fanciful in your alpinist ambitions. You decide to put your Himalaya plans on the back burner, intending instead to gain more experience on dry rock—scrambling, trad climbing, multi-pitch routes, and so on. This revision of conative attitudes represents a development in the particular way you are attached to alpinism, so that, now, your attachment partly consists in goals involving dry rock rather than the Himalaya. But now that your dry rock goals, and not your Himalayan ones, partly constitute your particular attachment to alpinism, you no longer have an attitude-dependent reason to await the winter, and instead have one to visit the Cairngorms in the summer. That is because satisfying the dry rock goal, not the Himalayan goal, is what it takes for you to engage appropriately with alpinism, given the concrete way your attachment to it has developed.

Whether or not our actions constitute valuable engagement with the things that matter can therefore very much depend on our conative attitudes, attitudes that we can form, revise, or abandon more or less at will. Applying the value-based theory of practical reasons to those cases where our actions do indeed constitute valuable engagement, the resulting attitude-dependent reasons are explained as *escapable* in the following sense. We can escape these reasons at will, that is, by simply changing the conative attitudes on which the value of what they say to do depend.

Not everything that is good to do depends in this way on our conative attitudes. Acting to preserve, protect, or at least not destroy what matters is always good, regardless of how our conative attitudes dispose us. Suppose, for example, you are finally realising your Himalaya plans, and find yourself stuck at base camp waiting out a blizzard. Ruminating in your wind-battered tent, you come to believe that you have bitten off far more than you can chew with the Himalaya, so you abort the plan entirely. With no attitude-dependent reason left to make the treacherous ascent, you retreat from base camp the next morning—for you, remaining there no longer amounts to valuable engagement with the Himalaya. But no matter your change of heart, no matter how little the Himalaya now figure in your conative attitudes, it will be good if you do not leave a mess of tarpaulin, empty oxygen canisters, and several weeks' worth of plastic food packaging at base camp. The value of respecting the Himalaya, and so our reason to do so, is *attitude-independent*.

Certainly, all reasons of respect are attitude-independent. By the same token, they are *inescapable* by simply changing our conative attitudes. A couple of remarks on the relevant sense of 'inescapability.' Firstly, contrary to its Kantian resonances, which associate it with what he called the 'categorical' status of some imperatives, the term 'inescapability' should not be taken to imply indefeasibility. As a tourist in the Himalaya, I have a reason to respect its natural beauty, which I only satisfy, say, if I do not disturb the wildflowers all around me in the foothills. But suppose another tourist will be gravely injured unless I run across the foothills to rescue him, which I cannot do without trampling many wildflowers underfoot. In that case, my reason to respect the natural beauty of my surroundings may well be defeated by my reason to prevent the fellow tourist from being gravely injured. But even so, I do not see what would force us to conclude that I thereby lost my reason to respect the natural beauty—that it was escapable after all. Secondly, it is not that we literally cannot escape our having reasons

of respect. This is perhaps a macabre point, but the dead have no normative reasons at all, and it is possible—sad but true—for people to take their own lives, and thereby escape their having reasons, *a fortiori* reasons of respect. So, the ‘inescapability’ of some reasons at will means: they are inescapable by simply changing our conative attitudes, at least on pain of death. That is the force with which reasons are ‘inescapable,’ in this sense, when they are.

4.3. Attachment-Dependent Obligations

No reason can at once be a reason of engagement and a reason of respect. That is because no reason can at once dependent and independent of our attachments. This mutual exclusivity, together with the fact that all attitude-dependent reasons are reasons of engagement, may lead us to think that all attitude-independent reasons—all inescapable reasons—must therefore be reasons of respect.¹ But this does not follow. This is clearest to see when we reflect on the nearby phenomenon of obligation. Doing so, I believe, in turn has repercussions for how we understand the force of ‘inescapability’ in the relevant sense.

Prima facie, the inescapability of reasons of respect looks like it may help to explain the phenomenon of obligation. Our obligations are often thought to be inescapable in something like the relevant sense, insofar as they strike us as binding independently of our goals, plans and desires, and thus as inescapable by simply changing our goals, plans

¹ This is a thought Samuel Scheffler attributes to Joseph Raz, when he says (2004: 261), “In short, the core of Raz’s proposal is as follows. Reasons for respecting values are categorical, whereas reasons for engaging with values are not.” There are certainly passages in Raz that suggest this view. For instance, he says (2001: 168) that for reasons of engagement, “the weight or stringency of these reasons depends on our tastes [...] or goals,” and continues, “Not so with reasons of respect: their stringency is not affected by our inclinations, tastes, goals, or desires.” But he also says (2001: 164) that reasons of respect are “more categorical” than reasons of engagement, suggesting that this is a graded notion, and thus that reasons of engagement may also be categorical, to a degree. This perhaps chimes with Raz’s (1977: 228) early suggestion that “all kinds of obligations [...] depend for their validity on the value of special bonds and many of them are [...] constituent elements of special human relations (husband-wife, parent-child, leader-led, etc.)” The view I am advancing can be taken in the spirit of the latter passages.

and desires. How to explain this characteristic feature of obligation? All reasons of respect, we said, are attitude-independent and therefore inescapable. A natural explanation, then, is that obligations are reasons of respect. This should not be taken as an equivalence. Quite apart from their inescapability, our obligations characteristically exclude certain countervailing considerations from guiding our actions, and reasons lacking this feature do not seem to obligate even if they are inescapable.² So, the natural explanation makes being a reason of respect, being inescapable, a necessary but not sufficient condition of being an obligation.

The explanation looks promising for many of our obligations. For instance, as it predicts, my obligation not to kill you does not seem to depend on my wanting or planning to keep you alive, or having that as a goal, or indeed any other conative attitude of mine. Even if I despise you and wish you dead, or I am indifferent to your fate, that does nothing to relieve me of my obligation not to kill you. The same goes for your obligation not to kill me. On the proposed explanation, that is because your life, my life, and human life more broadly, is precious. And like anything precious, we all have reason to respect it independently of our conative attitudes, which we do, at a minimum, by not destroying it. The same goes for our obligations regarding, for instance, the natural world. My obligation not to sully the Himalaya with non-degradable detritus does not seem to depend for its existence on my wanting, planning or aiming to keep it unspoiled, whether for love of its unique topography, or its mountaineering potential, or Nepalese culture, or any other conative attitude of mine. Likewise your obligation not to sully the Himalaya. That is because, the explanation goes, the sublimity of the natural world is a precious thing, and we all have reason to respect what is precious independently of how our conative attitudes dispose us to it. Thus, our obligations not to sully the Himalaya, and not to kill people, are inescapable by a simple change in our conative attitudes.

For many other obligations, however, their inescapability is not so readily explained in this way. The basic problem with these obligations is this: on the one hand, they seem to be inescapable in the relevant sense, but on the other, they are clearly attachment-dependent, and so bear the mark of reasons of engagement, not respect. Imagine, for

² As Raz says (1994: 40), “Duties are but a special kind of categorical reasons, they are peremptory reasons, reasons which exclude consideration of an prevail over certain categories of reasons.” See Chapter Five for more on the ‘peremptory’ (aka. ‘exclusionary’) character of obligations, that is, their requiring force.

example, that you have been hospitalised by a life-threatening illness. Other things equal, it is plausible that your mother has an obligation to be there for you at your bedside every night. Now that you have fallen severely ill, it would certainly be good if she did this. But for I, who do not know you, it would not obviously be good if I were to turn up at your bedside every night, or at the very least, not in the same way. This suggests that your mother's reason is a reason of engagement, not a reason of respect—one that depends for its existence on her attachment to you, in this case, as parent to child. And yet, we said, her reason is plausibly an obligation. This is plausible in part because it strikes us as attitude-independent, and therefore inescapable. To see this, suppose that your deep political disagreements had recently and explosively come to a head, with your mother admitting, sincerely, that she dislikes who you have become, and that she would rather not see you for a little while. These conative attitudes, sadly, are part of what give shape to the attachment between you and your mother as it has recently developed. Even so, it seems, your mother does not thereby escape her reason to be there at your bedside, and this is part of why her reason is plausibly an obligation. Of course, your mother's obligation, being attachment-dependent, cannot be a reason of respect. Such attachment-dependent obligations show that the inescapability of obligations is not in general explained by their being reasons of respect, as the natural explanation suggests.

Contrary to the natural explanation, we already know what in general explains why reasons are inescapable, when they are. Reasons are inescapable insofar as they are attitude-independent, which they are, on the value-based theory of practical reasons, because it would be good to do as they say independently of our conative attitudes. The value of respect is like this in general, though equally, as we have just seen, the value of engagement is not always not like this. So, it is not that actions are attitude-independently good to do only if they constitute respect for what matters. Both respect and engagement can be attitude-independently good to do, and when they are, *that* is what explains why our reasons to do so are inescapable, *a fortiori* when those reasons are obligations.

That said, the phenomenon of attachment-dependent obligation does nevertheless complicate the force of 'inescapability' in the sense at issue. Acting in satisfaction of our attachment-dependent obligations would constitute valuable engagement no matter our goals, plans, desires, etc., which is why, applying the value-based theory of practical

reasons, those obligations come out as inescapable by simply changing our goals, plans, desires, etc. In this, our attachment-dependent obligations are alike reasons of respect. But unlike reasons of respect, their existence is not entirely impervious to changes in our conative attitudes, because their existence depends on our attachments, and, as we saw, our conative attitudes partly constitute the history of the particular ways we have come to be attached to the things that matter.

To bring this out, let us revisit the case where we escape our attitude-dependent reasons by revising the conative attitudes in which our attachments partly consist. During a blizzard in the Himalaya, you decide to abort your ascent plans, leaving you no attitude-dependent reason not to retreat from base camp as soon as possible. Ordinarily, the old attitudes make way for the formation of new ones, in a process that gives shape to and advances the concrete history of our attachments. If you abort your Himalaya plans, that will usually not be because you have stopped making plans of an alpinist spirit altogether—perhaps you want to set your sights on a more attainable objective in the Himalaya or elsewhere, or prefer instead to focus on rock climbing, or fell-running, or whatever. But if you stop making plans or setting goals entirely, or if they are forever on hold without even a desire to rekindle them, then it seems, at the limit, that alpinism no longer matters to you—that your history of attachment to it has come to an end. Once you have so thoroughly neglected or become so totally estranged from this thing you used to love, it is no longer an object of your attachment, no longer a part of your life. And in losing the attachment, you lose any reasons that depend for their existence upon it, whether attitude-dependent or otherwise.

Such a possibility will have been palpable in the case where deep political disagreements with your mother led to a correspondingly deep shift in how she sees you, and how she is disposed to you in her conative attitudes. As the case was constructed, her new attitudes embody a weakening of the attachment between you, though they can only do this if they continue to partly constitute the attachment, albeit in less healthy shape than before. But we can easily imagine that, after you have recovered, your disagreements continue to flare up, leading you and your mother to become increasingly estranged. You figure positively in each other's conative attitudes less and less, with little desire for, or even an aversion to, reaching out by forming goals, plans or intentions actively involving one another. One consequence is that you have fewer and fewer attitude-dependent reasons regarding one another. But at the limit, if you truly stop

caring about each other, and are truly no longer part of each other's lives, then your attachment is over, and you lose even the attitude-independent reasons that depended on that attachment. Thus, in that event, were you to find yourself hospitalised once again, your mother (if you can call her that anymore) would have no attachment-dependent obligation to be there for you at your bedside, since the attachment on which it depends no longer exists. Other things equal, she has no more reason to do that than a stranger.

What this shows is that though attachment-dependent obligations cannot be escaped by *simply* changing our conative attitudes, they *can* be escaped by changing our conative attitudes in ways that erode our attachments over time, in the end destroying them entirely. This fact has implications for how to understand the force with which our reasons are 'inescapable,' in the sense at issue, when they are. To see this, consider again the inescapability of reasons of respect. 'Inescapability,' we said, should not be taken literally, because it is always possible in principle to take our own lives and thereby escape our reasons of respect. So, in the relevant sense, reasons are strictly speaking only 'inescapable' by a simple change of conative attitude on pain of death. But if attachment-dependent obligations are inescapable, they must be 'inescapable' in some other sense. For even if they cannot be escaped by a simple change in conative attitudes, they can, after all, be escaped by losing the attachments on which they depend. And so the force with which they are inescapable is not on pain of death, but only, less dramatically, on pain of losing an attachment. Conversely, we cannot escape our reasons of respect no matter how little we care for their objects—only our own death will achieve that. The worry is that this threatens our understanding of any general sense in which the reasons we want to say are 'inescapable,' are inescapable. For if there are both attachment-dependent and independent cases of attitude-independent reasons, then the sense in which they are 'inescapable' must differ, since the force with which they are inescapable differs.

4.4. The Force of Inescapability

Some reasons are inescapable on pain of losing our attachments, others are inescapable on pain of death. Is there some general way we can understand the force at play in both cases, so as to unify the sense in which reasons are 'inescapable,' when they are? What

could it be ‘on pain of’ that we cannot escape these reasons at will, which makes intelligible the loss of attachment and death as expressions of a single kind of eventuality?

I suggest that the sense of ‘inescapability’ is plausibly unified as follows. When reasons cannot be escaped at will, they are, in general, inescapable *on pain of losing the lives we are currently leading*. Of course, we lose the lives we are currently leading if we lose the one life that is ours to lead, that is, if we die. But we also lose the lives we are currently leading if we lose their essential constituents, that is, the attachments that structure, shape and give direction to the lives we are leading, in virtue of which we can engage with the things that matter and thereby lead our lives at all. The loss of attachments always marks a period of great upheaval in our lives, for their loss threatens the integrity, purpose and meaningfulness of the lives we are leading. This remains so even if the loss is incurred during the course of active self-transformation. The greatest upheaval of all, however, is the loss of life itself, which prevents the leading of a life of any quality, good or bad, by preventing the leading of life entirely. Such upheaval is what in general gives force to the inescapability of reasons at will, when they are inescapable. That is, in general, attitude-independent reasons are inescapable on pain of losing the lives we are currently leading.

This sense of ‘inescapability’ dovetails nicely with the explanation of attitude-independent reasons on the value-based theory of practical reasons. For there to be such reasons, according to that theory, there must be attitude-independent value in the actions they favour. Where the actions they favour aim not to destroy something, or even preserve and protect it, the actions only possess the value of respect if what is thereby not destroyed, preserved or protected itself matters. But if it does so matter, then respecting it would be good to do no matter how we are attitudinally disposed to it, making our reasons to do so inescapable—at least, that is, on pain of losing our lives, *a fortiori* the lives we are currently leading. By contrast, with attachment-dependent obligations, the attitude-independent value in what they say to do must be the value of engagement. For there to be such value in an action, I claim, not only must the object of our engagement be of value, but our attachment to that object must itself be of value. In my view, that is, the value of engaging with what matters cannot be attitude-independent unless the attachment our engagement depends on is itself of value.

Let me unpack these claims a little. Clearly, first of all, our actions cannot have the value of engagement if what we are engaging with does not matter. To see this, consider the following case, presented in a related context by John Rawls (1999: 379-80). Imagine someone who has a special affinity for counting blades of grass, and who has become attached to that practice, organising large swathes of their life around it. When such a person spends time mapping out an itinerary of parks, lawns and other green spaces for the coming weeks, or heads out to enjoy another fine day of meticulous blade-counting, do their actions possess the value of engagement? Presumably not, absent some special explanation. That is because the practice of counting blades of grass is pointless; it does not matter. And if engaging in that practice has no value, it has no value attitude-dependently or otherwise. So, by the value-based theory of practical reasons, the person has no reasons—attitude-dependent or otherwise—to engage in their unusual practice, however attached they may be to it.

I also claimed, however, that even if the object of attachment matters, so that there can be value in engaging with it, the resulting reason of engagement cannot be attitude-independent unless the attachment is itself of value. This claim may seem to be falsified by the existence of obligations that depend on attachments that are not obviously of value. I believe this objection stems from a failure to distinguish between the structure and the content that our attachments impart on our lives. This distinction is perhaps easiest to see with those of our attachments which do not take specific people as their objects, call them our *projects*. Our projects partly constitute the lives we are leading, in the sense that they form the structure or framework within which our engagement with their objects can influence how our lives are going in the first place, for better or for worse. As Katja M. Vogt puts it (2017: 129-30), “It is only via pursuits and activities of various kinds that one can aim to lead a good life.” But equally, our projects partly constitute the lives we lead in the quite different sense that they give them content, as determined by how we fare within the frameworks established by our projects. Success within those frameworks is an achievement, making our lives go the better for it; by the same token, failure makes our lives go worse. However, our projects could not make our lives go better or worse if they did not establish a framework for evaluating our engagement with their objects. Projects impart no content without structure. The same goes, though perhaps less clearly, for our attachments to specific people, that is, our *relationships*. Our relationships with specific others establish a framework within which

our interactions with them can be meaningful, which, depending on how we fare, makes our lives go better or worse to that extent.

In many cases where we are tempted to say that our attachments are no good, this is, I believe, because we are not doing very well by them, not succeeding within their parameters. This sort of thing happens all the time. We have all known people (perhaps ourselves) who are not very good at what they do, or who are not living up to a relationship very well. They have organised a large swathe of their lives around a project or a relationship which is filling their lives with failure, and all the pain and hurt that comes with it—these attachments seem no good for themselves or anyone else. I claim, however, that the value of attachments that is a condition of attitude-independent reasons of engagement obtains in virtue, not of the content, but the structure that our attachments impart. And unless our protagonists' unfortunate attachments were playing this structuring role, there would be no framework relative to which their engagement (or lack thereof) could be assessed as lacklustre. Thus, if their engagement is consistently lacklustre, so that the attachment on which that engagement depends seems no good, then their also continuing to have attachment-dependent reasons (obligations included) need not be a counterexample to my proposal. For the fact that their attachments seem no good, in that they are filling their lives with failure, does not necessarily exclude that their attachments are in fact good, in that they lend structure, shape and integrity to those very lives. It is the value of the structure, not the content, that explains the attitude-independent reasons.

I think the distinction between content and structure may be harder to see with relationships because poor quality engagement in relationships will eventually work to erode the structure they provide. The minimal way we can fail to properly engage in our relationships is when we engage no more than we are obligated to by our relationships, so that they become somehow flat or stale. This can have all kinds of causes, but whatever the cause, the ossification of our relationships shows up as a failure in the very lives they structure. But in continuing to structure our lives, such flat or stale relationships continue to support attitude-independent reasons to engage with the other, including obligations to revivify the relationship. Sometimes we feel we can and can be bothered to revivify such relationships, and sometimes not. Often the feeling reflects an assessment of the other person: what we think they are like, what makes them so, what really matters to them, and so on. Considerations as to what if anything attracts us to a

person are always relevant to whether we want to continue our relationship with them. But they make us apt to overlook the value of the relationship itself, our particular shared history. Whatever we think of the other, our failure to revivify the relationship is a failure to meet our relationship-dependent obligations. Trying to do so—involving, as it does, some degree of honesty about how we view one another—is a fraught business, with endemic risks of overreach, hurt feelings, and other interpersonal transgressions. Flat or stale relationships are in this way pregnant with the possibility of poor engagement of a more damaging kind, a kind that sours our relationships. Yet this possibility presupposes that soured relationships continue to impart structure, without which our attempts to make amends could have no impact on our lives, good or bad. At some point, however, poor engagement may leave our relationships so derelict, toxic, or indeed abusive, that they bring nothing but chaos and confusion to our lives, sapping away the sense of any activity within their sphere. Plausibly, we owe nothing to those with whom we are in an abusive relationship, and that, I submit, is because such relationships have lost their value: they no longer afford our lives a structure in which to lead them meaningfully. Short of that, however, our relationships will continue to give us obligations, even as we may be making a bit of a mess of them.

Why is the structuring role of our attachments crucial to explaining the obligations, and attitude-independent reasons more generally, that depend on them? For anything that matters, what it is for our attachment to play this structuring role *just is* that it would be good if we engaged with it in certain ways no matter how our conative attitudes dispose us to it. Our reasons to do so are in this sense ‘constitutive of’ the value of attachment, as is often said.³ So, if the attachment stops playing this structuring role, then if there remains value in engaging in those ways, the value cannot be attitude-independent. Consider again your mother’s relationship-dependent obligation to visit you in the hospital, despite her general lack of desire to see you following your dispute over politics. To explain her obligation, your relationship must still be an organising force in her life, or what is the same, it must be good if she were to visit you at the hospital whether she wants to or not. If the value in her visiting you could vary with her attitudes,

³ For instance, though his explanation differs, David Owens (2012) allows that some obligations constitute the value of being in certain relationships. According to Crescente Molina (2019: 16), Raz opposed any such view, though see Raz’ remarks at (1977: 228) for evidence to the contrary.

at will, your relationship could hardly be said to structure the life she is leading. Suppose that the political disagreement was in fact just a pretext for you to say hurtful, cruel and manipulative things to your mother, in yet another expression of a longstanding pattern of extreme and unprovoked emotional abuse. In that case, it may well not be attitude-independently good if she engaged with you, for example by visiting you in hospital.⁴ But even so, if she still wants to see you, and her desire to see you now partly constitutes your (toxic) relationship, then it could still be good if she engaged with you by going to the hospital. For all your toxicity, you do not cease to matter, and so the relationship can still give your mother reasons to engage with you. But only attitude-dependent reasons. Insofar as the only reasons she may have left to engage with you are attitude-dependent, your relationship, though it exists, is no longer an organising force in her life.

This explanation of attachment-dependent obligations nicely illustrates the force with which they are inescapable. Attachment-dependent obligations are not literally inescapable, since your mother, in the last case, manages to escape her relationship-dependent obligation to visit you in hospital. But she escapes her obligation only by becoming totally estranged from you, her own child, who she once cared for very much. This will be a great upheaval in her life, and not just owing to the abusive behaviour of the one from who she has become estranged—such upheaval is equally possible through a quieter, more passive dereliction of relationships. What matters is that she loses her valuable relationship to you, an essential constituent of the life she had been leading, an attachment in virtue of which that life had a particular structure, shape and direction it can no longer have. Those of our attitude-independent reasons, including obligations, that depend for their existence on our attachments are inescapable on pain of such loss. But so too are our attachment-independent ones, our reasons of respect, insofar as we

⁴ Here, my position diverges from John Gardner (2019: 223), who defends “the twin views that (i) barring exceptional cases, friendship is valuable and (ii) there are constitutive obligations (a.k.a. duties) of friendship.” The problem with Gardner’s view is what he has to say about the exceptional cases. If certain obligations are constitutive of relationships of certain kinds, regardless of the value of those relationships, then either we owe relationship-based obligations to those who abuse us in the context of relationships, or abusive relationships are in fact no relationship at all. One benefit of views like mine, on which obligations are constitutive not of attachments themselves, but the value of attachments, is that they can accommodate the existence of abusive relationships without conceding that the abused owe relationship-based obligations to their abusers.

can escape them too, but only on pain of death. For in our own death we lose the one life we have to lead, *a fortiori* the life we are currently leading.

4.5. Concluding Remarks

I have attempted to make headway on a sense in which reasons are ‘inescapable,’ when they are, by connecting some plausible ideas about the value our actions may have with the value-based theory of practical reasons. I distinguished the value of respect for what matters from the value of engagement with what matters, and correspondingly, reasons that do not depend for their existence on our attachments from those that do. The value of respect, and so reasons of respect, are also attitude-independent: they are inescapable by simply changing how we are disposed in our conative attitudes to the objects of respect, at least on pain of death. I argued, however, that the attitude-independence of reasons cannot be explained by their being reasons of respect, since our attachment-dependent obligations are attitude-independent reasons of engagement. Instead, the attitude-independence of reasons is in general explained by the attitude-independence of the value of the actions they favour. But since attitude-independent reasons of engagement are not inescapable on pain of death, this raises a question about the unity of the sense in which reasons are ‘inescapable,’ when they are. I suggested that the sense can be unified by interpreting the ‘inescapability’ of reasons as inescapability *on pain of losing the lives we are currently leading*. This interpretation, I argued, is supported by the value-based explanation of attitude-independent reasons of engagement, which presupposes, *inter alia*, that the attachments on which they depend are valuable, that is, that the lives we are currently leading would disintegrate without them.

In Defence of Second-Order Reasons

5.1. What are Second-Order Reasons?

I have been defending an account, fit to satisfy the soft anti-moralist desideratum from Part One, on which ordinary agent-regret acquires the desired ‘moral’ quality when it responds to failures to satisfy extraordinary reasons for action, namely, our obligations. In the previous chapter, I defended an account of the inescapability of obligation. In this chapter, I aim to bolster accounts of the requiring force of obligation that appeal to the concept of *second-order reasons*.

What exactly do I mean by ‘second-order’ reasons? We can begin to get the idea on the table as follows. In the theory of practical reasoning, it is generally accepted that considerations that have a normative bearing on action do so either positively or negatively. That is, they speak either *for* or *against* that action; they are the *pros* and *cons* of doing it. Reasons for action must in this sense be either positive or negative reasons for action.¹

It is also plausible that considerations can have a normative bearing on action by bearing on doing it in certain ways. The fact that dinner is in the fridge might be a reason for going to the kitchen, but the fact that the baby is asleep in the kitchen might be a reason not to do so noisily. Situations like this are ordinary enough, and the description

¹ Bare instances of ‘reasons’ refer to normative reasons unless otherwise specified.

seems to be faithful to their normative structure. It seems right that the fact that the baby is asleep in the kitchen figures as a reason not to go there noisily, as opposed to a reason not to go there full stop. After all, there may be other ways of going to the kitchen that will not wake the baby, such as, for example: quietly.

Doing something quietly is, or at least can be, one way of doing it amongst many others. Some believe that doing it for a certain reason can be another of those ways.² If they are correct, it follows that considerations can have a normative bearing on action by bearing on doing it for certain reasons. Considerations like this are what philosophers have in mind when they speak of *second-order reasons*. Like all reasons for action, they think, second-order reasons are either positive or negative: they are either reasons for acting for certain reasons, or reasons against acting for certain reasons.

The idea of second-order reasons was introduced to Anglophone philosophy by Joseph Raz in his classic 1975 book *Practical Reason and Norms*. One of the major projects of that book is to showcase the theoretical benefits of recognising second-order reasons, by showing how reasons with requiring force—decisions, norms, rules and obligations—cannot be adequately explained without appealing to negative second-order reasons, or as Raz says, ‘exclusionary’ reasons. Roughly put, the line of thought is as follows. What intuitively unifies the relevant phenomena is the special stringency they claim for themselves as reasons for action. We are required to do as they say, in that we cannot escape doing so just because we do not want to, or because it would inconvenience us, or even because we have other genuinely more worthwhile things to be getting on with. The special stringency of these reasons seems to lie in how they take matters out of our hands by excluding certain options from deliberation independently of their merits. It is this exclusionary force, a *sui generis* ability to defeat reasons to do otherwise irrespective of comparative weight, which is supposed not to be adequately explicable until we recognise second-order reasons. But once we do, the thought goes, we can say that the relevant phenomena defeat certain reasons to do otherwise in this *sui generis* way because they are negative second-order reasons, in this case, reasons not to act for those reasons to do otherwise.³

² To do something ‘for’ a reason is to do it guided by that reason.

³ Raz’s explanation of the *sui generis* mode of ‘exclusionary’ defeat, which he offers in his postscript to the second edition of the book (1999: 188-90), tends to get overlooked even by the friends of exclusionary reasons. The idea, in brief, is to invoke an independently plausible

Many philosophers have agreed with Raz on the need for second-order reasons to explain these normative phenomena (and others like them), allowing them to play a role in their various theories.⁴ But others have been more resistant to second-order reasons, seeking to challenge their legitimacy.⁵ The challenges fall into two broad groups. *Theoretical* challenges say that the phenomena that allegedly cannot be explained without second-order reasons are either not in fact explained by them adequately, or amenable to explanation using less contested concepts.⁶ But to mount a theoretical challenge is already in a way a concessive move, since it grants that second-order reasons are coherent in the first place. *Conceptual* challenges target precisely this assumption. There are indefinitely many ways to do this, in principle, but the dominant strategy has

principle that governs conflicts between reasons. Suppose there are two conflicting reasons, *A* and *B*, where *A* is weightier than *B*. Conforming with *A* completely precludes conforming with *B*, whereas conforming with *B* need not completely preclude conforming with *A* (e.g. because some other action constitutes partial conformity with *A*). In that case, we can conform with both reasons to some degree so long as we conform with *B*. Plausibly, there is a principle governing such conflicts that at least sometimes directs us to do whatever allows for the greatest measure of conformity with the applicable reasons, which in this case rules out conformity with the weightier reason *A*. Raz's thought is that the conflict between exclusionary and excluded reasons is like this. Since conforming with exclusionary reasons only requires that we not act for the excluded reasons, we can conform with both exclusionary and excluded reasons just so long as we conform with the excluded reasons without being guided by them. Thus the principle may apply, and if it does, then it requires us to do whatever will allow for conformity with both reasons. That is, it rules out acting for the excluded reason, which it can do, we have seen, even if it is the weightier reason.

⁴ Big name examples from various traditions would include John Gardner and Timothy Macklem (2004), Raimo Tuomela (2012), Anna Stilz (2013) and Cass Sunstein (2016).

⁵ After an initial wave of challenges, followed by responses from Raz, there have been one or two subsequent waves that have gone largely without comment from Raz or anyone else. Of the initial wave, probably Michael Moore (1989), Stephen Perry (1989), Donald Regan (1989), Larry Alexander (1990), Heidi Hurd (1991) and Frederick Schauer (1991) are the most searching. Subsequent notable efforts include Emran Mian (2002), Christian Piller (2006), Stephen Darwall (2010), Daniel Whiting (2017) and Noam Gur (2018). Points raised by Moore, Perry and Regan are addressed directly in a response piece by Raz (1989), as are those raised by Darwall in Raz (2010). Many of the challenges have been addressed (directly or indirectly) or anticipated by either the response pieces, Raz's postscript (1990), or an unpublished article uploaded by Raz (2021) to SSRN.

⁶ For a recent conspectus of both negative and positive theoretical challenges, as well as a novel positive theoretical challenge, see, respectively, parts 1, 2 and 3 of Gur (2018).

been to try to put conceptual pressure on the idea that we can act for a reason for a reason.⁷ We will explore the dominant strategy in more detail in what follows.

In this chapter I focus on conceptual challenges to second-order reasons, and in particular on the robust conceptual challenge advanced by Daniel Whiting (2017).⁸ In section 2, I show how Whiting's challenge instantiates the dominant strategy and present his CREDIT ARGUMENT in the wider context of that strategy. I go on in sections 3-5 to argue that the CREDIT ARGUMENT has shaky premisses, and that however charitably they are interpreted, the argument comes out either unsound or with a dramatically weakened conclusion. I conclude in section 6 that Whiting's challenge is unsuccessful, removing a serious barrier to accepting the Razian account of requirement.

5.2. Understanding the Conceptual Challenge

The dominant strategy aims to put conceptual pressure on the idea that we can act for a reason for a reason.⁹ Instances of the strategy can be thought of as having basically two components: first, a *principle* that licenses the inference from the existence of second-order reasons to our ability to act for a reason for a reason, and second, an *argument* that establishes the incoherence of our possessing such an ability. In this section I bring out this bipartite structure in Whiting's challenge. Subsection 5.2.1 spells out the principle, and subsection 5.2.2 the argument.

⁷ We first find the dominant strategy in Moore (1989: 875-83). I believe that the reply in Raz (1989: 1174-8) is enough to disarm Moore's challenge, so I make no attempt to relitigate that particular dispute here.

⁸ At the time of writing, and to the best of my knowledge, the only discussion of Whiting's challenge is in Nate P. Adams (2021), where its success is taken at face value as a partial basis for rejecting, or at least substantially revising, Raz's view of second-order—and thus exclusionary—reasons.

⁹ Here, and in what follows, I take the referents of words like 'acts' and variables like 'φ' to include omissions. I also assume that reasons against actions can be understood as reasons for the corresponding omissions. This is common in the literature, though cf. John Gardner and Heike Jung (1991) and Justin Snedegar (2018). Expressions like 'reasons to act for reasons' and their cognates are thus used to say things about second-order reasons in general (i.e. both positive and negative) without littering the text with parenthetical qualifiers.

5.2.1. Principle (LR)

The principle component may differ in content for different instances of the dominant strategy, but its general form, its gist, can be expressed as the following schematic claim about reasons for action.

(R) If p is a reason for S to φ , then it is possible for S to φ for p .

Something like schema (R) looks highly plausible. It captures the widespread intuition that it is in the nature of reasons for action that they can guide our action.¹⁰ If schema (R) is true, and we grant the existence of second-order reasons, it entails that we can act for a reason for a reason. To see this, consider that, in general, a second-order reason takes the form of a reason for S to ψ -for- q . Hence, if we make p in schema (R) a second-order reason by plugging ψ -for- q into the slot for φ , then it will follow, if (R) is true, that it is possible for S to ψ -for- q for p , i.e. that S can act for a reason, q , for a reason, p .

We can specify schema (R) in different ways depending on how we read ‘possibility’ in its consequent, and many specifications state sensible looking principles governing reasons for action. But proponents of the dominant strategy are constrained in their choice of principle by the bipartite structure of their strategy. If the preferred principle gives too trivial a reading to ‘possibility,’ then even if it is true, it will be tricky to make the argument that it is not possible, in the same sense, to act for a reason for a reason. This is why metaphysical or physical readings of ‘possibility’ in schema (R) are nonstarters. It is not obvious how one would go about arguing that acting for a second-order reason is metaphysically impossible in the way, say, that giving birth to one’s own father seems to be. Nor does proving the physical impossibility of acting for a second-order reason seem any more a task for philosophy than does proving, say, that matter cannot travel faster than light. But too committed a reading of ‘possibility’ leaves the resultant principle at risk of losing its veneer of truth. In that case, it becomes too easy to contest whether the conclusion of the argument component of the strategy speaks to

¹⁰ See for example Korsgaard (2008a: 31) and Raz (2011: 28).

the coherence of second-order reasons at all, even if the argument itself is compelling.¹¹

Whiting's preferred specification of schema (R) sits in a sweet spot between these extremes: committed enough to enable the construction of an argument, but not so committed as to be controversial. He calls it the *motivational constraint* (MC) on reasons, which at first glance looks the same as schema (R): "If that p is a reason for you to ϕ , it is possible for you to ϕ for the reason that p " (2017: 403). But Whiting goes on to specify a reading for 'possibility' in (MC). First of all, he says that "To endorse (MC) is not to endorse the controversial 'internal' view of reasons" (ibid.). He then adds that "To endorse (MC) is not to endorse the controversial view that there is an 'epistemic filter' on reasons" (ibid.). Putting these qualifications together, we can say that for it to be relevantly possible for S to ϕ for p , S need not know that p , nor need S 's subjective motivational set be such that they would or could conclude that they ought to ϕ by a sound deliberative route, were they to know that p .¹²

'Possibility' in Whiting's (MC) is thus very weak. To see this, suppose that S is completely ignorant of the facts, that they are entirely devoid of motivation, and that their doing ϕ would not anyway be conducive to realising whatever value there may be in satisfying p . Even in that case, it could still be relevantly possible for S to ϕ for p . Given the existence of a person S and the fact that p , there is no way the world could contingently be—i.e. as regards S , the wider facts, or their relations—that would rule out that it is possible for S to ϕ for p in the sense at hand. To highlight this point, we will call this a purely 'logical' sense. 'Possibility' in Whiting's (MC) means *logical possibility* in

¹¹ This is in effect the trouble with Moore's version of the dominant strategy. Moore (1989: 875) endorses a Hartian specification of schema (R), where 'possibility' is read in terms of having both the capacity and a fair opportunity to act for a reason. What this means, for Moore, is that the existence of second-order reasons would entail that those to whom they apply are able to rationally choose the reasons for which they act, which he argues is inconsistent with the concept of causation. When Raz replies that second-order reasons entail only rational control over the reasons for which we act, as opposed to full-blown rational choice, he is contesting the Hartian specification of schema (R), or at least Moore's interpretation of it. This is why Raz can agree with the conclusion of Moore's argument, as he does when he writes (1989: 1177), "we cannot choose but we can control."

¹² This is not to say that for it to be relevantly possible for S to ϕ for p , *no one*—not even an ideal agent—need have the relevant knowledge or subjective motivational set, since this would make (MC) highly implausible. Only that the one to whom the reason p applies, viz. S , need not have it.

just this sense. In other words, his preferred specification of schema (R) is more perspicuously expressed as follows.

(LR) If p is a reason for S to ϕ , then it is *logically* possible for S to ϕ for p .¹³

Principle (LR) may be resisted due to the obscurity of ‘logical’ possibility. Admittedly, it is hard to get a grip on the sense in which it is ‘possible’ for S to ϕ for p , such that the possibility obtains independently of how things could contingently be with the relevant bits of the world. To do so, it will be helpful to reflect on the different ways it can be self-defeating for S to try to ϕ for p . In particular, I claim, it is logically possible for S to ϕ for p if and only if it is not necessarily self-defeating for S to try to ϕ for p . We can illustrate what it means for a course of action to be necessarily self-defeating with the following example.

Ben plays saxophone in an improv group. He has high ambitions for audience experience, always looking to deliver a powerful moment to remember. That it would lend spontaneity to the performance may seem to be a reason for him not to try, in performing, to construct the powerful moment he is after. There really would be value in the performance being spontaneous: the musical result would be less derivative, more ecstatic, and ultimately more powerful. But if Ben plays with one eye on this enhanced musical potency, always acting on this basis, then what he will be doing is precisely: trying to construct a powerful moment. Aspects of the value of musical performance have put Ben in a bind. It is not possible for Ben not to try to construct a powerful musical moment if he acts on the basis that it would lend spontaneity to his performance. For if he acts on the basis of this consideration, he *necessarily* fails to do as it would have him do, namely: not to try to construct a powerful moment. There is no way things could contingently be with the relevant bits of the world that would make it possible for Ben to do as the consideration would have him do on its basis. In this sense, the bind is a logical one. It is logically self-undermining for Ben to try to ϕ for p , and thus logically impossible for him to ϕ for p . This, together with (LR), is why the fact that it would lend spontaneity to his performance is not after all a reason for Ben not to try to construct a

¹³ This principle is considered in a perceptive paper by Gardner and Macklem (2012: 455-7), to which the following discussion is greatly indebted.

powerful moment, even though the spontaneity that would be achieved in so doing really would enhance the musical results. Whatever reasons Ben may have not to try to construct a powerful moment, this is not one of them.

When it is self-defeating for S to try to ϕ for p , it need not be necessarily so. It is often the case that p is a reason for S to ϕ , but due to contingent features of S or the practical context, it will actually impede S 's satisfaction of p if they set out to ϕ by being guided by p . Consider the distorting effects a job market known to be hyper-competitive can have on the psychologies of entrants to it. A prospective interviewee of course has a reason to give a solid interview, viz. that it could get them job security. But for them to actually try to do so with job security on the mind (how wonderful it would be to be free of applications, to save for the future, to make rent) is liable, the market known to be such as it is, to whip up a state of frenzied over-preparation, which will in the end be detrimental to their giving a solid interview. Though it happens to be self-defeating for the interviewee to try to give a solid interview guided by their job security reason, here it is only *contingently* so. It is not necessarily self-defeating because there are ways the world could be such that trying is actually conducive to succeeding, as would be the case, for instance, were the job market not known to be in such a dire state.

With the notion of logical possibility clarified, I suspect that many will find principle (LR) uncontroversial. For it is an extremely weak principle, albeit one whose application is often overlooked.¹⁴ It does not rule out that we have reasons for action even when we are completely ignorant of them, or unmotivated, or indeed when it would be contingently self-defeating to try to act on them. It is also not inconsistent with alternative specifications of (R) that put stronger necessary conditions on when a fact counts as a reason. Given that (LR) is not implausible and in any case very weak, I grant it as an assumption in what follows.¹⁵

¹⁴ Moore (1989: 875) rejects (LR), as it applies to moral reasons, on account of this very weakness. He reasons that any specification of principle (R), insofar as, and because, it applies to moral reasons, must be a stronger 'substantive' principle (e.g. grounded in the morality of fairness) rather than a weakly 'formal' one. I do not see why this should be so. Whatever else moral reasons are, they are a species of reasons, and so if the existence of reasons is in general constrained by a formal principle, it seems *ad hoc* to suggest without argument that things should be any different for moral reasons.

¹⁵ I do not suggest that (LR) is totally unassailable. For instance, Gardner and Jung (1991: 571) have argued that there is a fundamental asymmetry between reasons for and against action,

5.2.2. The CREDIT ARGUMENT

To the extent that Whiting's challenge instantiates the dominant strategy, we can expect there to be an argument as to why the idea that we could act for a reason for a reason is incoherent. Principle (LR) dictates what exactly it is that needs to be argued for here. The argument needs to establish that we cannot for a reason for a reason in the sense that it is not logically possible for us to do so.

To this end, Whiting (2017: 406) offers the following line of thought.

A plausible view is that, if you ϕ for a good reason, that is, for an undefeated reason which favours ϕ ing, you are creditworthy [...]. Suppose that Kelly decides to go to the pub for the reason that she promised Dave to meet him there [...]. Here, Kelly does what she ought to do and deserves credit for this.

Now suppose that Kelly decides to go to the pub for the reason that she promised Dave to meet him there for the reason that she will get a reward if she acts in a creditworthy fashion. That Kelly will be rewarded is a second-order reason—a reason for her to act for certain first-order reasons. However, in this case Kelly is surely not creditworthy for going to the pub.

The case in which Kelly acts for a reason for the reason that she will get a reward shows that it cannot be true *both* that acting for a good reason suffices for creditworthiness *and* that it is possible to act for a reason for a reason. So, we must reject either the claim about credit or the supposition that in this case Kelly acts for a first-order reason for a second-order reason.

I suggest that we reject the latter.

We have here the argument Whiting takes to complete his conceptual challenge. I

namely, that whilst we can act for reasons for action, we cannot act for reasons against action. If this asymmetry holds, then assuming that reasons against action are equivalent to reasons for omissions, (LR) could be questioned in its generality across both reasons for and against action.

believe it is faithfully reproduced in the form of a valid *reductio ad absurdum*. We can call it the CREDIT ARGUMENT.

The CREDIT ARGUMENT

P1 When *S* ϕ s for the right reason, they are creditworthy.¹⁶

P2 It is logically possible for *S* to ϕ for a reason for a reason.

P3 But when *S* ϕ s for the right reason for a yet further reason, they are not creditworthy.

C Thus, when *S* acts for such further reasons, they both are, and are not, creditworthy.

Since the conclusion is a contradiction and follows from *P1*-3, we are compelled to reject one or more of these premisses. Whiting's suggestion is that, faced with these options, we are compelled in particular to reject *P2*. Needless to say, we are compelled to reject *P2* only on the assumption that it is unacceptable to reject *P1* or *P3*. Since the CREDIT ARGUMENT is a valid argument, I will, in the rest of the chapter, inquire into whether it is sound: whether it can safely be assumed that neither *P1* nor *P3* can be rejected. Sections 5.3 and 5.4 focus on *P1*, and section 5.5 on *P3*.

5.3. The Right Reasons are Second-Order

I want to start by observing that the 'right reason,' as it appears in *P1*, admits of first- and second-order readings. From the perspective of Whiting's challenge the correct reading must be first-order, else the CREDIT ARGUMENT will assume what it aims to prove incoherent. The problem, I will suggest, is that *P1* is only plausible on its second-order reading. To see this, we first need to unpack the basic thought that *P1* encodes, in virtue

¹⁶ I say 'right' reason where Whiting says 'good,' though nothing hinges on the choice of words. What matters is that they are both intended, minimally, to describe the reasons acting for which we are intuitively creditworthy. For the record, I prefer my word for this because it is standard in the literature, including the literature Whiting uses to bolster *P1*. I therefore continue to use it. The narrower reading of 'good' available in the quoted passage, where the 'good' reasons are the undefeated ones, is discussed in detail in section 5.4.

of which it is plausible, and then to show how it already invokes second-order reasons.

There is a powerful strand in moral philosophy whose central thought is that moral worth accrues to those who do the right thing for the right reasons.¹⁷ The plausible intuition driving the thought is that it is in some sense not enough for an action to manifest moral worth that it happens to be the right thing to do. It must moreover be motivated in the right way, responsively to what is morally relevant in the situation at hand. This is the basic thought behind *P1*, and it has an impressive pedigree. In the *Nicomachean Ethics* (2014: 9.8 1169a18-22), for example, Aristotle says that a morally serious person would sacrifice their life for the sake of their friends or country out of clear-eyed pursuit of the noble. Or more amenably to the modern ear, perhaps, there is Kant's suggestion in the *Groundwork* (1996: 4:397-8) that the shopkeeper who charges a fair price from the motive of duty therein manifests a good will, unlike one who does so merely to preserve their reputation. Let us call the basic thought behind *P1* the *Aristotle-Kant* thought.

For our purposes, it is helpful to consider the Aristotle-Kant thought as it bears on contexts of choice. We can separate *what* is chosen from *how* what is chosen gets chosen. The emphasis on motivation in the Aristotle-Kant thought concerns less the 'what' than the 'how' of choice. But the proper significance of how we choose what we choose will escape us unless we first understand what it is that we choose from: our options. Suppose a shopkeeper charges £2 for a pack of toilet roll, and that they do so because that is a fair price. Should we say that charging £2 is the option the shopkeeper chose? In one familiar sense, the answer is yes. But that sense, though familiar, is misleading. For in that sense, the shopkeeper chose from an infinity of price options: n options to charge £ n . Of course, for many n (e.g. $n = 10$ billion), to charge that price would be totally absurd.¹⁸ To make

¹⁷ This is not to say that its representatives agree on the finer details of what the thought amounts to. Most influential in the contemporary literature are probably Nomy Arpaly (2002) and Julia Markovits (2010). It should be noted that Whiting (2017: 406) acknowledges the strength of this literature, and indeed seems to take this all by itself to push the rejection of *P1* beyond the pale. I agree that this is an excellent literature, and that it lends serious support to *P1*. As we will see, however, I do not agree with Whiting that it necessarily supports *P1* on his preferred first-order reading.

¹⁸ Hence, Raz (2011: 178) lambasts as "silly" the suggestion "that as I write I have a choice not only between carrying on with this chapter or taking a break to listen to the news, but also

any money at all, as shopkeepers are wont to do, it would be pointless to charge £10 billion for a pack of toilet roll. In a less misleading sense of ‘option,’ that is not really an option. Thus, our options are not just actions, but actions done for reasons: to charge £2 because it is fair, or £5 because it capitalises on a supply shortage, and so on.¹⁹ In Korsgaard’s (2008b: 218) colourful phrase, we choose “the whole package.” What we choose when we choose to do the ‘right thing’ is, like any option, a whole package: an action done for a certain reason. The significance of how we choose what we choose lies in what it reveals about the value we see in the relevant whole package, the value we see in doing a certain thing for a certain reason.

Put this way, we see that the Aristotle-Kant thought is consistent with different substantive views on the content of this value. It is in this sense a metaethical thought. The content of this thought, I submit, is that moral worth is a matter of acting for a certain type of positive second-order reason. To see which type, we need to consider the ways in which, when there is value in doing something—call this the *basic* value—there may in turn be *further* value in doing it guided by recognition of its basic value. Since there being basic value in doing something is a reason, *p*, to do it, the ways there may be further value in doing it guided by its basic value are so many ways there can be reasons, *q*, to do it for the reason *p*, i.e. so many types of positive second-order reason.

The figure 1 table may be helpful as a visual aid to the proposed typology. There are four axiological scenarios to consider, which vary along two dimensions: the relations between the basic value and the further value, and the relations between acting from recognition of the basic value and the realisation of the further value. There are,

between either of these options and cutting off my finger or my ear, or taking my shoes out of the cupboard and putting them back in again, etc.”

¹⁹ What about those cases some believe to be felicitously described as cases of choosing an option for literally no reason, such as the nice example in Ruth Chang (2004: 82) of choosing one can of soup from an aisle of identical cans of soup? I must admit to not finding it entirely clear how these cases are meant to work. At one level of description, the can *is* chosen for the same reason that any can would have been, viz. that one is responsible for dinner tonight, or something like that. If it is said instead that the object of choice is not *this* can, but *this-rather-than-that* can, then it is not clear to me that the context should be described as one of choice. Or no more so than, say, my ‘choice’ this morning to shave in whatever fine-grained pattern of strokes I did, as opposed to any other (suitably non-eccentric) pattern.

correspondingly, four types of positive second-order reason.²⁰

<i>In what relation does acting from recognition of the basic value stand to the realisation of the further value?</i>		
<i>In what relation do the basic and further values stand?</i>		
	Contingent	Constitutive
Sameness	TYPE A	TYPE C
Difference	TYPE B	TYPE D

Figure 1. Proposed typology of positive second-order reasons.

In Type A cases, the further value is the very same value as the basic value. For example, if Fred needs food, then there is value in his going to the shop insofar as this satisfies his need, which gives him, other things equal, a reason to go to the shop. Maybe he ends up there because he goes past it whilst jogging, or because he has to collect a package, or because he ducks in to avoid the rain. In each case, Fred goes to the shop, thus satisfying the reason he has to go to the shop in virtue of his need. Equally, he could go because he sees that doing so has the value of satisfying his need. Like any of the ways that Fred could end up at the shop, ending up there guided by his need for food would of course satisfy that very need. That is, the further value of going to the shop guided by the basic value in doing so is none other than the basic value. As we saw in subsection 5.2.1, it can be a contingent matter whether or not the basic value of a given course of action also doubles as a further value. In Fred's case, it does, just so long as his going to the shop for his need-based reason remains an effective way of going to the shop. But if this ceases to be, then it becomes contingently self-defeating for Fred to go to the shop for his need-based reason.²¹ In that case, if there remains any further value in Fred's going to the shop guided by the basic value in doing so (viz. need-satisfaction), it cannot

²⁰ Sometimes I speak in terms of values and the reasons they 'underwrite,' 'provide' or 'supply,' etc. As far as I can tell, the validity of the typology that follows does not depend on the metaphysical priority of values over reasons this may be taken to express.

²¹ Situations where the basic value in someone's doing something does not double as a further value in their doing it guided by the basic value, such that doing the latter is contingently self-defeating, are a precondition for a Razian authority justifiably to step in according to his 'normal justification thesis' (1986: 53).

be the same as this basic value.

Type B cases differ in that the further value is a different value from the basic value. For example, if it would please Fred's therapist were he to make an effort with his day-to-day, and his going to the shop guided by the value in doing so (viz. need-satisfaction) would prove this, then on the one hand, that it satisfies his need is a reason for him to go to the shop, and on the other, that his doing so for that reason would please his therapist is a reason for him to do so for that reason. This is a second-order reason to go to the shop for the need-based reason, but it differs from—and competes with—the Type A one Fred also possesses. We know it differs because the further value that underwrites it is not the value of satisfying Fred's need, but the quite different value of pleasing his therapist. The basic and further values come apart. As with Type A cases, the further value in Type B cases provides a reason to act guided by the basic value only insofar as doing so is an effective way to realise the further value, which is a status it has contingently. Perhaps what pleases Fred's therapist depends on her mood: sometimes she advises that being kind to oneself consists in upholding day-to-day tasks, and other times that it requires relaxing the demands one makes of oneself day-to-day. In the latter moods, Fred's going to the shop for the need-based reason will no longer please his therapist, although it will, other things equal, still facilitate the satisfaction of his need for food. Fred's going to the shop for the need-based reason will then only be supported by the Type A second-order reason, not the Type B one.

What unites cases of Type A and B is that the further values can be realised independently of whether the protagonists act from a recognition of the basic values. Hence why, in these cases, acting from such recognition may or may not realise the further value, contingent upon whether or not so doing actually stands in the appropriate instrumental relation to realising the further value. When it does, the second-order reason there is to do so is a *contingent* one. This contrasts with Type C and D cases, where the further value cannot exist independently of action guided by the basic value. In these cases, the relation between acting guided by the basic value and realising the further value is *constitutive* rather than contingent. The difference between Type C and D cases turns on whether or not the basic and further values are one and the same. Plausibly, if there can be *sui generis* value in declarations of true love, then realising it requires satisfying a Type C constitutive second-order reason. The *sui generis* value, we are assuming, cannot be realised without the declaration, but there is no such

declaration unless it is made for the reason that one truly loves its recipient. Similar forms of words or gestures produced for any other reason cannot but fail to realise the *sui generis* value, which is what it means to say they are empty. With Type D cases, by contrast, the basic and further values are different (as in Type B cases), but still the further value cannot be realised except by acting from recognition of the basic value in doing so (as in Type C cases). We will consider examples of Type D cases in a moment. For now, suffice it to say that the further value in both Type C and D cases cannot be realised without guidance by the basic value, which explains why they provide constitutive positive second-order reasons.

Typical examples of morally worthy behaviour marshalled by proponents of the Aristotle-Kant thought emphasise their Type D character, often by contrasting them with Type B variants where moral worth is intuitively lacking. Consider Markovits's (2010: 227-8) cases of the reward-seeker and the altruist. Both rush into a burning building from a recognition of the fact that doing so enables the rescue of a trapped child. But they each see a different further value in doing so guided by that same basic value. For the reward-seeker, the further value is that she will thereby earn a reward. This value accrues to her rushing into the building guided by the basic value of the child's life only insofar as a reward happens to be set up for her doing so, as would be the case, for instance, if a millionaire with insight into the reasons for which she acts offers her such a reward. Here the further value underwrites only a Type B contingent second-order reason. For the altruist, by contrast, the further value lies in due concern for the intrinsic value of the child's life, or something like that. Rushing into the building in order to rescue the child is constitutive of what will manifest such concern, such that this value, given the situation and its concrete demands, simply cannot be realised in any other way. The further value the altruist sees supplies a Type D constitutive second-order reason. Needless to say, it is the altruist and not the reward-seeker whose action is supposed to be morally worthy.

A similar structure can be discerned in Arpaly's (2002: 227-30) cases of Ron the extremist and Huck Finn. Ron wants to kill Tamara, but refrains so as to realise the value (read: 'value') of not killing members of his own ethnic group, to which Tamara also belongs. It is clear that Ron's acting guided by this basic 'value' stands only contingently related to the further value of protecting Tamara's life. Had the world been set up differently—if Tamara happened, say, not to share Ron's ethnicity—then Ron's acting

from the basic ‘value’ he does would no longer be an effective way of preserving Tamar’s life. We thus have a contingent second-order reason of Type B. By contrast, when Huck lies to the slave hunters in order to save Jim, he does so because he sees that this is what the further value of due respect for Jim’s humanity demands, however viscerally, inexplicitly and against his own conscience. Upon being confronted by the slave hunters, the situation’s concrete demands leave Huck without a way to realise this value other than by lying to them in order to save Jim. Here the further value supplies a constitutive second-order reason of Type D. Though Ron and Huck both do the right thing, it is Huck alone that Arpaly wants to suggest acts in a morally worthy way.

For our purposes, Markovits, Arpaly and other proponents of the Aristotle-Kant thought need not share any substantive view on the further values that determine moral worth in their preferred examples, nor need they be happy with our way of parsing the examples out. All that matters is that their common Type D structure can be discerned. What unites proponents of the Aristotle-Kant thought is that the further value in their preferred examples, whatever its content, supplies *constitutive positive second-order reasons of Type D*. Given that *P1* is plausible only insofar as it expresses this distinctive metaethical thought, the ‘right reasons’ should be given a second-order reading.

5.4. The Right Reasons are not First-Order

I suggested that the sorts of examples typically used to support *P1* are best described in second-order terms, as cases of Type D. But that suggestion is, after all, just a suggestion, and I have not tried to argue that the examples cannot be described in first-order terms. So, perhaps charity demands that we give the ‘right reasons’ in *P1* a first-order reading. The most sensible proposal along these lines is probably Whiting’s (2017: 406) own view that the first-order reasons acting for which is creditworthy are the *undefeated* reasons. Is there more to be said for the CREDIT ARGUMENT on this reading?

I do not think so. Acting for an undefeated reason is simply not sufficient to make the action creditworthy. This can be established with counterexamples involving very mundane contexts of action, where whatever we do there is nothing especially significant at stake. Take the choice at the breakfast bar between apple juice and orange juice. This is probably a choice between incommensurables, such that neither reason for

drinking a juice is defeated by the other: both are undefeated. But it is implausible that we would therefore deserve credit for drinking either juice for the reason in its favour. This is partly due to the implication that we deserve credit whatever we do, which we may, and not implausibly, think is ruled analytically out by the very notion of creditworthiness. More basically, however, what is implausible is that we could deserve credit for something so utterly mundane, just because we do it for an undefeated reason. We can see this by tweaking the incommensurability out of the case, so that there is a clearly superior option, for instance, by making the choice one between a normal glass of apple juice and a tiny one. Other things equal, our reason to opt for the tiny glass is defeated by our reason to opt for the normal one, so that only the latter reason is undefeated. But even if we opt for the normal glass guided by the undefeated reason, so that we are justified in doing so, it is again implausible that we are thereby creditworthy for so doing. As Gardner (2007a: 104) says, “more is required for credit than is required for mere justification.”²²

This shows that for *PI* to come out true with the ‘right reasons’ read as undefeated reasons, it needs to be modified. The difficulty lies in modifying it such that it comes out true, whilst retaining the link it seeks to establish between the right reasons and creditworthiness. The only real candidate seems to be something like this:

*PI** When *S* ϕ s for the right reason (i.e. an undefeated reason), they may be, but need not be, creditworthy.

This is probably the strongest true proposition that connects acting for the right reason and creditworthiness, reading the ‘right reasons’ as undefeated reasons. We could arrive at *PI** by observing that whilst it is not plausible that acting for an undefeated reason is sufficient for that action’s being creditworthy, it is plausible that an action’s being creditworthy is sufficient for its being done for an undefeated reason. So when someone acts for an undefeated reason, it is perfectly possible that they deserve credit for this. That is what *PI** captures.

²² I remain silent on the question whether it is also the case that *less* is required, i.e. whether it is not only not sufficient but also not necessary for creditworthiness that an action be justified. If that is so, it favours the view that, unlike justification, people are principally creditworthy for something other than what they do (e.g. what they intend to do, who they are, etc.).

Modifying *P1* to *P1** may well restore soundness to the CREDIT ARGUMENT, but it does so by blocking its desired conclusion. If, as *P1** says, *S* may, but need not be creditworthy in ϕ ing for the right reason, then there may, but need not follow a contradiction in the presence of the other premisses. For only some of the second-order reasons mentioned in *P3* will generate a contradiction: those that bear on acting for whichever ‘right reasons’ it is creditworthy to act for. Only *those* second-order reasons turn out to be incoherent. But that is a much narrower claim than Whiting hoped to establish. The desired conclusion is that second-order reasons are *all of them* incoherent, but on the modified CREDIT ARGUMENT, some of them turn out not to be.

5.5. Do the Examples Generalise?

Perhaps the foregoing discussion leaves the reader quite cold. Or perhaps the reader agrees that reading the ‘right reasons’ in terms of undefeated reasons is no good, but is willing to hold out for *some* more promising first-order reading. After all, nothing in the intuitive resonance of the Aristotle-Kant thought makes it such that the failure of the undefeated reasons reading forces us to accept a second-order one. So let us grant that some first-order reading of *P1* is acceptable. In this final section, I argue that this does not help the CREDIT ARGUMENT, since *P3* is false.

P3 says: when *S* ϕ s for the right reason for a yet further reason, they are not creditworthy. One way to bring out the falsity of *P3* is to point out that it is based on examples from Whiting the features of which simply do not generalise. The first of these (2017: 406) we have already seen:

suppose that Kelly decides to go to the pub for the reason that she promised Dave to meet him there for the reason that she will get a reward if she acts in a creditworthy fashion. That Kelly will be rewarded is a second-order reason—a reason for her to act for certain first-order reasons. However, in this case Kelly is surely not creditworthy for going to the pub.

Call this example *Pub*. In *Pub*, the fact that Kelly promised Dave to meet him at the pub is a first-order reason for her to go to the pub, and the fact that Kelly will get a

reward if she acts in a creditworthy way is a second-order reason for her to go to the pub for the first-order reason that she promised Dave to meet him there. Incidentally, Whiting seems aware of the risk that his examples might not generalise, since he offers the second example to convince the reader “that nothing turns on the particular example” (ibid.).

Whiting’s second example (2017: 406-7) is as follows:

Kelly is deciding whether to send her daughter to school A or school B. The career-related considerations favour B, but the weightier education-related considerations favour A. Kelly promised Dave that she would make her decision on educational grounds alone. So, Kelly decides to send her daughter to A for educational reasons on the basis of her promise to Dave. As it happens, had Kelly not made her promise, she would have sent her daughter to B. She is concerned about her daughter’s education only because she is concerned with keeping her promise to Dave. That is, she responds to the education-related considerations only for the reason that she promised Dave to do so.²³

Call this example *School*. There are various distracting factors at play in *School*, but its core content is as follows. That her daughter will receive a better education at the educationally better school is a first-order reason, an education-related reason, for Kelly to send her daughter there (in this case school A). That she promised Dave to make her decision only on the basis of education-related reasons is a second-order reason for Kelly to send her daughter to school A for a certain first-order reason, namely, the education-related reason.

The first-order reasons in *Pub* and *School* are meant to strike us as the ‘right reasons,’ if any are. Let us grant that these first-order reasons are indeed the right reasons. In that case, *Pub* and *School* really do bear out the truth of *P3*. After all, in both examples, Kelly is described as acting for the right reason for a further reason, whilst our intuition tells us that she is not creditworthy for acting as she does. Nevertheless, I submit, we can agree on intuitions around *Pub* and *School* without swallowing any general claim about

²³ Whiting (2017: 399) imagines the education-related reasons to be “That the school is nearby” and “that it offers philosophy,” and the career-related reasons to be “that having a child there would further Kelly’s career.”

the creditworthiness or otherwise of acting for the right reason for a further reason, such as is expressed in *P3*. The intuition that Kelly is not creditworthy for acting for the right reason in *Pub* and *School* can be attributed to features of the examples other than that she does so for a second-order reason. Consider the following variants of *Pub* and *School*.

*Pub** Kelly goes to the pub for the reason that she promised Dave to meet him there, out of respect for the practice of promising.

*School** Kelly decides to send her daughter to the educationally better school for the reason that she will receive a better education there, out of love for her daughter.

In *Pub**, Kelly has the same first-order reason to go to the pub as in *Pub*, viz. that she promised Dave to meet him there. But her second-order reason to go to the pub for this first-order reason is a different one, and could be rendered as the fact that going to the pub for the reason that she promised Dave to meet him there is what would constitute the value of respect for the practice of promising in the circumstances at hand. Similarly, in example *School**, Kelly has the same first-order reason to send her daughter to the educationally better school as in *School*, viz. that she will receive a better education there. But again, her second-order reason to send her daughter to the educationally better school for this first-order reason is a different one. The relevant consideration is that sending her daughter to the educationally better school for education-related reasons is what properly constitutes her love for her daughter.

In terms of the typology of section 5.3, the basic difference between the two pairs of examples is that the positive second-order reasons in *Pub* and *School* are plainly *contingent*, whereas in *Pub** and *School** they are more plausibly *constitutive*. In *Pub* and *School*, it is a contingent matter that the further values accrue to Kelly's acting for the first-order reasons. Were the world instead set up so that the reward in *Pub* were not earnable by Kelly's acting for her promissory reason, or so that her promise to Dave in *School* were not fulfillable by acting for her education-related reason, then acting for these first-order reasons would stop being an effective way to realise the relevant further values, and the second-order reasons they underwrite would evaporate. But in *Pub**, it seems that, given her promise to meet Dave at the pub, it is constitutive of what it takes

for Kelly to properly respect the promising practice that she meet Dave at the pub for the reason that she promised.²⁴ Similarly, in *School** it seems that, now that her daughter is about to start school, Kelly simply cannot make good on her love for her daughter except by making the schooling decision for education-related reasons alone. If my proposed typology is anything to go by, then it should be no surprise that Kelly is intuitively creditworthy in *Pub** and *School**, but not in their unstarred counterparts.

Whatever the merits of the typology, the key point for the present argument is that *Pub** and *School** differ from their unstarred counterparts only in terms of the second-order reasons. The first-order reasons are untouched, and presumably, therefore, remain the ‘right reasons.’ Like their unstarred counterparts, then, *Pub** and *School** are cases where Kelly acts for the right reason for a further reason. Yet unlike in their unstarred counterparts, Kelly is intuitively creditworthy in *Pub** and *School**. So, if there is an intuition that Kelly is not creditworthy in the unstarred cases, what their starred counterparts show is that it stems from some feature other than that they involve acting for the right reason for a yet further reason. Someone’s acting for the right reason for a further reason is insufficient to rule out their being creditworthy in so doing. *P3*, in other words, is false.

Faced with cases like *Pub** and *School**, defenders of the CREDIT ARGUMENT could try to amend *P3* to accommodate them in the following, somewhat *ad hoc*, sort of way.

*P3** When *S* ϕ s for the right reason for an *irrelevant* further reason, they are not creditworthy.

The idea with *P3** is to stipulatively define the ‘irrelevant’ further reasons as the class of second-order reason found in *Pub* and *School*, and similar such cases, whatever class that may be. Perhaps it is extensionally equivalent to the class of contingent positive second-order reasons found under Type A, Type B, or both—though it need not be. Cases like *Pub** and *School** would not be counterexamples to *P3**, since the protagonists in those cases, though creditworthy, are stipulated to act for the right reason for ‘relevant’ further reasons. So, unlike *P3*, *P3** does not claim that the protagonists in *Pub** and

²⁴ There can be other ways to show respect for the promising practice (e.g. by running a campaign, or making a documentary), but the point is that, for Kelly, given her promise to Dave, she now cannot do so other than by doing as she promised for the reason that she promised.

*School** are not creditworthy. Such cases give us no reason to think *P3** is false.

Once again, the problem with this manoeuvre is that the contradiction it generates will be limited in scope as per the stipulation: it is only when people act for *irrelevant* further reasons that they come out both creditworthy and not creditworthy. In turn, the CREDIT ARGUMENT will only establish that *some* second-order reasons are incoherent, namely, the irrelevant ones. And this is, once again, a much narrower claim than desired. Whiting's challenge set out to show second-order reasons incoherent *tout court*, but it now admits that, in fact, some second-order reasons *are* conceptually coherent: the *relevant* ones.

5.6. Concluding Remarks

The CREDIT ARGUMENT aims to expose second-order reasons as incoherent. I proposed a novel typology of second-order reasons, suggesting that *P1* of the CREDIT ARGUMENT is plausible only if we read the 'right reason' as a certain type of second-order reason. I then argued that on the alternative first-order reading, *P1* comes out false, a fault that can only be rectified by seriously weakening the argument. Finally, I argued that even if some other first-order reading turns out to be workable, this will not save the CREDIT ARGUMENT, because *P3* is anyway false, and is, again, not salvageable without seriously weakening the argument.

I conclude that the CREDIT ARGUMENT does not go through, and that Whiting's conceptual challenge collapses with it. As ever, important questions remain, such as those surrounding the distinction between constitutive positive second-order reasons of Types C and D, and whether there are yet other versions of the dominant strategy that merit further study. For now, I would be happy to have said enough to show that Whiting's own version is unsuccessful. In so doing, I hope to have rendered the concept of second-order reasons more serviceable to those attracted to explaining normative phenomena in its terms, and thus to have removed another serious obstacle to accepting the Razian account of requiring force.

Thus concludes the chapter, and with it, Part Two of the dissertation. Let me briefly zoom out to the larger picture. In Part One I argued that if soft anti-moralism is to be a viable way of avoiding the excesses of both moralist and hard anti-moralist responses to

tragedy, we need to account for a feeling of responsibility that is: (a) apt without fault, and (b) of a sufficiently 'moral' quality. The aim of Part Two has to vindicate the account outlined in the preamble, which is, in a nutshell, that *ordinary agent-regrets in response to failures to satisfy extraordinary reasons* fit the bill. Chapter Three defended the claim that agent-regret is a fitting response to failures to satisfy reasons for action *at all*, and Chapters Four and Five addressed worries about the inescapability and requiring force of obligations, which is what I suggest makes them so extraordinary as reasons. Though my aim has been to address the deepest objections to the account, I suspect there will remain some resistance to it, even if I have managed to make it somewhat more plausible than before. Nevertheless, I hope, at least, to have entitled myself to the account in Part Three, where my focus shifts in a political direction.

Part Three

Politics

So far, I have aimed to answer the question: how should we each respond to contingency in our lives? I have concerned myself especially with perhaps the most troubling face of contingency in our lives, which I am calling ‘tragedy.’ Tragedy befalls us, in the relevant sense, when we fail to live up to our important obligations even though we were not at fault—our failure was, in that sense, beyond our control. We are left apt to suffer feelings of moral responsibility, despite our lack of fault in the situation.

Some seek to moralistically deny the coherence of tragedy, by insisting on the highly intuitive Fault Principle, which makes fault a condition on apt feelings of responsibility as typically understood, that is, on blameworthiness. Others, impressed by the powerful phenomenology of tragic cases, instead deny the Fault Principle, leading to a ‘hard’ form of anti-moralism. My aim in Part One was to situate and assess this debate. Chapter One situated it as an instance of a more general dilemma about our responses to contingency in life, and in particular the phenomenon of moral luck. In Chapter Two, I argued that

we can only escape the dilemma by ‘softening’ our anti-moralism, but that the viability of doing so requires an account of a novel feeling of responsibility that is: (a) of a ‘moral’ quality sufficient to do justice to the phenomenology in tragic cases, and (b) capable of aptness in the absence of fault.

In Part Two, I outlined and defended an account of a novel feeling of responsibility suited to play this role, in terms of ordinary agent-regret in response to failures to satisfy extraordinary reasons for action. Chapter Three defended the claim that ordinary agent-regret is in general an apt response to failing to satisfy reasons for action, extraordinary or otherwise. In Chapters Four and Five, I went on to explain the extraordinariness of some reasons for action in terms of their inescapability and requiring force, that is, in terms of their status as obligations. Our failure to satisfy important obligations leaves us apt to feel agent-regrets whose special ‘moral’ quality lies in the difficulty of justifiably letting go of or repressing them. This is how I claim we should each respond to tragedy, a uniquely troubling face of contingency in our lives.

I believe that this makes soft anti-moralism viable, and with it, an escape from the moral luck dilemma as it arises for tragic cases. We may, however, worry that we have only achieved this by fiat. Perhaps, that is, our so-called solution amounts to little more than reasoning as follows: ‘The Fault Principle is a condition on blameworthiness; agent-regret is no form of blame; therefore, problem solved.’ But is it any less anathema to commonsense intuitions about the ethical significance of control that we can be fittingly plagued by agent-regret for breaching important obligations, even if it was beyond our control? Part Three attempts to make plausible the following reply. Although we do not have *personal* control over whether tragedy befalls us, we do, nevertheless, collectively enjoy a measure of *political* control over where and upon whom it tends to fall in society. In a slogan, my reply is: ‘Tragedy has a politics.’

I will offer philosophical analysis of two real-world normative phenomena, both of which suggest, in different ways, that we do indeed have political control over where and upon whom the burdens of tragedy tend to fall in society. We exhibit this control, I claim, in the politics of our social relations: in struggles over the demands that in fact constitute our social relations, and over how we collectively understand their value for human life. In Chapter Six, ‘Contractualisation and its Discontents,’ I address a phenomenon of the latter kind, wherein changing perceptions about our social relations threaten to render us, in a certain way, disposable to one another. In Chapter Seven, ‘Security, Social

Relations, and the ‘Political,’ I address the former possibility, focusing on how the ways we practice the norms of our social relations necessarily influences who is most exposed to the risk of tragedy in the social order, and at whose expense.

By analysing these phenomena, I mean, however gesturally, to address the question: how should we, together, respond to contingency in our lives? I am suggesting that our collective response to tragedy should be to engage in its politics—if, as I claim, they are in fact there to be engaged with. This is not, however, a claim for which the reader should expect any argument in Part Three. Rather, my idea is this. So long as the analyses offered in Chapter Six and Seven are coherent, then *if* they also strike us as apt to the phenomena, we will have no reason to doubt that the phenomena exemplify the politics of tragedy. The proof of the slogan, ‘Tragedy has a politics,’ will be very much in the pudding. I leave it for the reader to judge whether or not there is enough in Part Three to appease our intuitions about the ethical significance of control.

Contractualisation and its Discontents

6.1. What is Contractualisation?

As indicated in the preamble, my goal in Part Three is to advance plausible analyses of instances of tragedy that bring out its political dimensions, thereby making plausible the slogan: ‘Tragedy has a politics.’ To this end, in this chapter, I consider the phenomenon that some have called the *contractualisation* of our social relations. Before anything else, we should get clear on what is to be understood by this term.

As I use the term, ‘contractualisation’ should be taken to mean no more and no less than what the philosopher and legal theorist John Gardner says he means by it. In his words (2018b: 36),

The social process of contractualisation, as I will call it, is the process by which contractual norms come, not only to be socially supported, but to enjoy a unique pre-eminence amongst those that enjoy social support. What kind of pre-eminence? This kind: whenever the norms of certain roles and relationships can be understood as having their source in a contract, they are so understood.

To give a flavour of the kind of thing he means, he continues (2018b: 37),

Police officers, social workers, architects, and solicitors typically carry out their professional work under a contract, whether it be a contract with an authority, a firm, or a client. [...] However, the professional obligations of police officers, social workers, architects, and solicitors are not contractual obligations. They do not have their source in this or any other contract. [...] Such roles have their own normative structure and no amount of contracting can change it. The contract may get some or all of the professional obligations wrong, and may even impose conflicting obligations.

Clearly, for Gardner, part of the problem of contractualisation is epistemic. As our social relations become increasingly contractualised, we increasingly misrepresent both the source and content of the obligations having which partly constitutes our standing in those relations. But the “creeping contractualisation” (ibid.) that Gardner so clearly regards as pernicious is not only an epistemic malaise, but an ethical one too. Once our social relations become contractualised, Gardner thinks, their demands are no longer “supposed to contribute constitutively” to our lives but “only instrumentally,” coming to be “conceived as an alien force consuming what would *otherwise* be [our lives]” (2018b: 45). Contractualisation tends to occlude from view the distinctive value potential in our social relations, rendering us unfree to engage in them in a proper, unalienated, fashion. The irony, Gardner thinks, is that the political and ideological forces behind the ‘creep’ are liberalising forces in some sense (the ‘neo-’ one perhaps), making contractualisation “a freedom-destroying monster with a freedom-friendly face” (2018b: 46).

I agree with Gardner’s worries about the creep of contractualisation, though my emphasis will differ. The ills of contractualisation that concern me have to do with how it makes it intelligible that we treat social relations as disposable and, correspondingly, how easily it allows interpersonal failures to go unaccounted for—including failures we incur faultlessly. In this way, I suggest, the contractualisation of social relations disrupts our moral-psychological practices in response to faultless failure, making tragedy in our social relations sensitive to the political and ideological forces behind contractualisation. Not only does contractualisation affect the extent to which we regard agent-regret as apt on the part of those who fail faultlessly, the extent to which we see ourselves as exposed

to tragedy by our social relations. It also affects the extent to which we see our relations as worth salvaging, or whether reconciliation is in order, leaving those who are let down exposed to yet further failure.

The material in this chapter covers a fair distance, but its structure, I believe, should be simple enough to follow. The first section, section 6.2, makes a *positive* contribution. There, I analyse the ills of contractualisation, explaining how the phenomenon exhibits the politics of tragedy. In 6.2.1, I present the contractual relationship as an instrumental one, which makes the value of meeting its constitutive demands—the contribution that doing so makes to our lives—itself instrumental. In 6.2.2 I then use this instrumentalist view of contract to explain how contractualisation makes it intelligible that we treat our social relations as disposable, and how this relates to tragedy. Section 6.3 then proceeds *negatively*, mounting an extended case against those who deny my central claim that the value of contract is first and foremost instrumental. In 6.3.1, I deal with a range of views that may helpfully be brought under the heading of what historians of ideas call the *doux commerce* thesis. I then address, in 6.3.2, a perhaps more sophisticated style of proposal, one rooted in the distinctive late-modern value of *personal detachment*. In section 6.4, I tentatively conclude that none of the non-instrumentalist alternatives look like they can be made to work, thus removing the main obstacle to my positive analysis of the ills of contractualisation.

6.2. A Moral-Philosophical Analysis

Contractualisation affects how we see our social relations. Their constitutive obligations erroneously come to be seen as having their source, not in the special non-instrumental significance of the social relations they partly constitute, but instead in the value of some contractual arrangement between the parties. An analysis along these lines presupposes a favourable account of the value of contractual relations. I argue, in 6.2.1, that we should take contracts at face value: the main contribution they make to our lives is instrumental. In 6.2.2, I use this instrumentalist view of contract to analyse contractualisation and its ethical consequences, foregrounding the way it renders our social relations disposable. I try to show how this disposability influences our perceptions about where the burdens of tragedy fall in the wake of interpersonal failure.

6.2.1. Contract as an Instrumental Relationship

Contracts are made between contracting parties. Those parties have a material existence, and so contracts must do too. We are thus familiar with smart contracts and pdf files, as well as more rustic forms like paper documents or cuneiform tablets. We can even, at the limit, make do with a mere pattern of grunts and gestures shared between the parties. But it is not sufficient for the contract to exist that the raw materials do. It must have, in addition to a material existence, what we could call a normative existence. Contracts are a source of *obligations*. In what follows, we focus on *performance* obligations: obligations to perform as contracted.

Contracting parties do not owe it to each other to do anything until they have made a contract between them. Performance obligations are thus relationship-dependent, that is, they do not exist without a contract. As with attachment-dependent obligations more generally, it can be tricky to explain our performance obligations. The difficulty, we saw in Chapter Four, lies in how to explain the sense in which they are ‘inescapable,’ if, given their attachment-dependence, they may be escaped by shedding the relevant attachment. I think we should resolve the difficulty for performance obligations in much the same way as we did, in Chapter Four, for attachment-dependent obligations in general. Their ‘inescapability’ at will should be taken as inescapability *on pain of losing the lives we are currently leading*. I believe this resolves the difficulty, but the resolution is only available on the assumption that the attachments on which our performance obligations depend are themselves of value. In other words, to explain our performance obligations, it needs to be the case that the contractual relations on which they depend are of value in some way. So, the question is: what, if anything, is the value of our contractual relations? What kind of contribution do they make to our lives? In the contract literature, this is a hotly contested question. But not, I do not think—this is of course tendentious—because it is such a difficult one. It is, I believe, one of those rare cases where the most obvious answer also happens to be the correct one. Let me explain.

Many of our social relations stand to benefit us directly. They admit of the kind of non-instrumental engagement that, all being well, will thereby enrich our lives. Plausible examples would include friendships and familial relations, as well as (suitably fleshed out) the sorts of civic and professional relations on Gardner’s list: police officers, social

workers, architects, solicitors, and the like. With such social relations, I claim, we know what it is—even if it is hard to articulate—to be enriched by engagement with them for their own sake. To my mind, it is very clear that contractual relationships do not belong in this category. Our lives do not obviously stand to be enriched directly by engaging in contractual relations. I am not sure what it would even mean to engage with them ‘for their own sake.’ In the usual case, I take it, the point of our entering into a contract is *to do what for each of us is in our best interest, when that happens to be something we cannot do without the other*. You have a reason to do what is in your best interest; I have one for mine. When it happens to be the case that neither of us can satisfy our respective reasons without the performance of the other, that is where contracts come into their own. It is hard to deny or overstate the value of the contractual relationships that result. But that value, it seems to me, is instrumental value *par excellence*.

To reiterate: I believe the value of contract is straightforwardly instrumental. We each have a reason to do whatever is respectively in our best interest, and if it so happens that neither of us can do those things without the performance of the other, the contract finds its point. That will remain so, as long as: (a) our respective best interests continue to provide a reason to do those things, and (b) we each continue to be unable to do those things without the performance of the other party. Of course, our interests are constantly twisting, turning and evolving in response to situational changes, broadly construed to include changes both outer and inner. What our best interests give us reason to do has a correspondingly malleable nature. It all depends on the situation: our goals, skills or ambition may change, opportunities or obstacles may come and go, technology may unlock more efficient ways of doing things, or the political calculus may shift. This is no doubt a partial list. But any of them, in any combination, could affect our best interests such that (a) or (b) no longer obtain. In either case, I claim, the contract becomes devoid of the only value it was obviously meant to have: instrumental value.

We saw, in Chapter Four, that we cannot explain the obligations constitutive of our social relations unless those relations are of value in the sense that they play a structuring role in our lives. Tokens of those types of relations are thus relevantly of value so long as they are not so derelict, toxic or abusive as to bring chaos and disorder into our lives. Should they stop being non-instrumentally valuable in this way, they no longer explain the associated obligations. I suggest that the same goes for social relations whose value is instrumental, like our contractual relations. However, the value of our relationships is

much more fickle when that value is instrumental, rather than non-instrumental. Take friendship, for example—a paradigm non-instrumentally valuable relationship. When friends confront such situational changes as above, that does nothing, *ceteris paribus*, to make their friendship any less a source of non-instrumental value for them. The value of the relationship tolerates such situational changes, and thus with it, the relationship-dependent reasons that value explains. By contrast, with a contractual relationship, the very same situational changes may alone suffice to undermine the instrumental value of the relationship, by negating conditions (a) or (b). In that event, we thereby stop having the dependent reasons. Our performance obligations are in this way correspondingly fickle.

It may seem implausible that our performance obligations lapse the moment our contractual relationships stop being instrumentally valuable. After all, the law of breach of contract recognises obligations on the part of parties to pay compensatory damages for non-performance even in the case of efficient breach. Breach becomes efficient precisely where the contract loses its instrumental value for the non-performing party. We may wonder: if performance obligations obtain in virtue of the instrumental value of the contractual relationship, then why does the law say that we have compensatory obligations even where it recognises that that value ceases to obtain?

This legal data does indeed speak *prima facie* against the proposed instrumentalist view of contract. However, a closer look, I suggest, will reveal that my proposal not only accommodates the legal data, but in fact explains the expectation measure of contractual damages that is standard in common law jurisdictions. The basic idea is that the state may have its own reasons to establish legal obligations that do not track the extra-legal obligations that constitute our social relations. Contracts as I have described them are social relations whose demands have a grip on us quite apart from there being a coercive state hovering around to recognise and enforce them.¹ Once we see this, there is room to say that our legal obligations to compensate in cases of efficient breach reflect reasons

¹ This is a controversial view, though it is one I hold in the good company of Gardner. In fact, according to Gardner, the law itself can agree with our view. He suggests (2018b: 33-4) that this is demonstrated by the existence of legal protections for trade unions to facilitate legally non-binding contracts between workers and bosses via collective bargaining, as was the case in Britain before Thatcher. To deny that extra-legal contracts can exist would, moreover, have the unusual effect of making much of the social contract tradition something of a misnomer, with the important exception, perhaps, of Hobbes.

of the state, rather than those constitutive of our contractual relationships.

Imagine two strangers find themselves stranded on a remote island. They each want to have a dip in the sea, but to do so, they need to leave their belongings on the beach, where they are at risk of loss—perhaps by the wind, a bird, or a person. It would be in the best interest of each if they could each enlist the services of the other to mind their belongings as they take turns in the sea. They each have a reason to go for a dip, though neither can do so safely without the contribution of the other. The circumstances are ripe for the parties to enter into a contractual relationship, which, as things stand, would be of instrumental value to each of them. Of course, going into the relationship, they both know that the situation may change for the other party with respect to conditions (a) and (b), in which case the instrumental value of the relationship for the other will lapse, and with it, their reason to perform as agreed. What they need is to give the other party a reason not to act for whatever reasons a situational change might give them for non-performance. It is not obvious how they could do so—or not, at least, without in effect renegotiating to less advantageous terms—except by means of a *threat*. Here on the remote island, the posture of the parties is: ‘If you do not do your bit, I will make you do it.’

Whether such a threat is enough to do the work of normative reinforcement is, for better or worse, contingent on the properties of the contracting parties. Are the parties evenly matched in the relevant physical attributes, e.g. speed, strength, stamina? What about psychological attributes, e.g. willpower, courage, mercilessness? Or—for a slightly less remote island perhaps—social attributes, e.g. reputation, connections? Depending on how these attributes appear, the parties may, in light of situational changes, calculate that the benefits of non-performance defeat the costs, that is, those associated with the realisation of the other’s threat plus any sunk costs accrued along the way. They may, in other words, calculate that breach will be efficient. In that event, if the other calculates that the gains of realising their threat defeat whatever costs attend doing so, the result will be a pathetic descent into some form of interpersonal violence.

Thus, I claim, there is a certain volatility built into extra-legal contractual relations. On the one hand, they enable us to do what is in our best interest, but on the other hand, they only do so by invoking the ugly mechanism of fear and threats of violence. These are the Janus faces of our contractual relations. We may instinctively want to object on the grounds that the proposal lends contracts an unpalatable Hobbesian flavour. But we

ought not to let moral judgment suppress those features of practical reality it finds unpalatable. On the remote island, it plainly matters that the strangers pose enough of a threat to one another that neither party seems easily taken advantage of. To see this, consider that if one party were obviously much more vulnerable to a power play, then any agreement between the parties would not count for very much. Were a vulnerable party to reason in this way, I do not see that they exhibit a viciously Hobbesian cast of mind.

Let us now take leave of the remote island and return to civilisation. The first thing to note is that parties to contract will be no less parties to contract. They cannot enjoy the instrumental benefits of contracts without running the risk of interpersonal violence between them. But here in civilisation their interactions take place against the backdrop of the state. The state, I claim, has a standing reason to support at least some contractual relations between its citizens such that they can enjoy their distinctive benefits without risking a descent into violence, which would, of course, be a moral and societal calamity. It achieves this, I claim, by taking the enforcement threatened by the contracting parties out of their hands and placing it into the hands of the courts. This manoeuvre keeps the original mechanism of normative enforcement more or less intact. It is just that the threat is now posed by a (hopefully) legitimate source of violence, namely, the coercive apparatus of the state. Our legal obligations to pay compensatory damages in the wake of efficient breach reflect these reasons of state, as opposed to the reasons constitutive of our contractual relations themselves, which, by hypothesis, have lost their instrumental value as per conditions (a) and (b).²

² This tracks, and, I think, partly explains the fact that the law *prima facie* recognises only those agreements where the parties can be presumed to intend to establish legal relations. The doctrine of intent to contract enters English law with the appellate decision in *Balfour v Balfour* [1919] 2 KB 571, the gist of which is that agreements within intimate relationships cannot be presumed to be intended to be legally enforceable, even where consideration is present. The idea is that agreements made within such relationships are typically upheld by trust rather than threats, and so cannot be presumed to be contractual. On my view, the state has no standing reason to enforce such extra-legal agreements, since threats are not endemic to their operation. By contrast, commercial agreements are typically made between strangers, and so must—given the absence of trust (at least in the matter of the agreement)—be contractual in nature. Thus, on my view, the state has a standing reason to recognise commercial agreements as contracts, and to enforce them as such. I am suggesting that in English law this is achieved by the *Balfour*

Before closing, let me explain, as promised, how the proposed instrumentalist view of contract is able to vindicate the expectation measure of damages that is standard in the common law of contract.³ Let us return to the remote island. Imagine that having entered into the contract, the situation changes such that the best interests of one party, *D*, would now be better served by non-performance. If there is still instrumental value in the contract for *D*, the only thing making it so is the threat of their counterparty, *P*. I rendered the threat as something like: ‘If you do not perform, I will make you.’ Of course, that *D* performs does not matter non-instrumentally to *P*. What matters to *P* is whatever instrumental contribution *D*’s performance makes to *P*’s satisfying their reason to swim in the sea without losing their belongings, which was the whole point of their entering into contractual relations with *D* in the first place. This does not change, I take it, in the event that *D* decides to breach. It is not that *D*’s (now late) performance matters to *P* non-instrumentally, such that only specific performance—or something as close to that in spirit as is now possible—would count as satisfactory by way of compensation. What matters to *P* is to extract from *D* whatever would bring *P* closest to where they expected to be, given the agreement, in terms of satisfying their reason to swim in the sea without loss. It may be that something like specific performance fits the bill. But it may equally be something else: a quantity of island currency, perhaps, sufficient to recoup any lost belongings, or to buy someone else’s services, or whatever, as the case may be. In short, *P*’s threat is to extract the cost of *D*’s non-performance in terms of the instrumental value they expected to reap by contracting in the first place. In this way, I claim, an expectation measure of damages is installed deep in the architecture of the proposed instrumentalist account.⁴

doctrine, which establishes the rebuttable legal presumption that commercial parties do in fact intend to invoke the coercive machinery of the courts.

³ This is the view, for example, of the House of Lords in *Co-operative Insurance Ltd v Argyll Holdings Ltd* [1998] AC1 (HL) 11, where specific performance is said to be “an exceptional remedy.” It is often remarked that the exception and the rule are reversed in civil law systems, where it is specific performance that enjoys primacy in some sense or other. This may be true in theory, but it is far from the case in practice, as is persuasively shown by Ronald J. Scalise Jr. (2007: 730-3) in a helpful summary of the comparative legal scholarship on the issue.

⁴ As Robert Cooter and Melvin Eisenberg put the point, “expectation damages place on the promisor the promisee’s loss of his share of the contract’s value in the event of breach, and thereby sweep that loss into the promisor’s calculus of self-interest” (1985: 1463).

6.2.2. Disposable Social Relations

For a given type of social relation, contractualisation sets in as the constitutive demands of its tokens are taken to have their source, not in the type of relation that it is, but in a contractual arrangement between the parties. This encourages, as we saw Gardner note, various epistemic errors. To start with, the performance obligations that are supposed to formalise the demands of the social relations therein contractualised may get the idea of those relations wrong, “by garbling the associated obligations” (2018b: 37). This will come with ethical repercussions: it sets up incorrect normative expectations, leading to misguided efforts at best, and interpersonal failures at worst.

But suppose we were to assume away the errors of contractualisation as regards the contents of obligation. Even still, we would be prone to making a further epistemic error as regards their source, with all the ethical ramifications that error entails. The error stems from the fact that the affected social relations are of a non-instrumentally valuable type, whilst the value of contractual relations, on the proposed view, is instrumental. Once those social relations have been contractualised, they are no longer to be regarded as sources of meaning in our lives, to be engaged with for their own sake. Instead they are to be regarded as subordinate to the pursuit of some further end, ends we may have quite independently from the existence of the relevant relations. Contractualisation in this way warps our understanding of the contribution our social relations make to our lives.

To echo Gardner again, this is a process that we should collectively accept only with extreme caution. After all, it describes our alienation from forms of social relations that are plausibly constituents of the good life. To the extent that our values are dependent on underlying practices—on constellations of attitudes and behaviours—the process of contractualisation will, if anything, tend to erode our ability to participate in the social-relational values at issue. That being said, until we have ceased to practice these values—until we reach what we could call ‘total’ contractualisation—it is *their* demands that we ought to be responsive to, *not* their shadowy contractualised counterparts. I would like to focus on the ills that this process inflicts upon our social relations *en route*, as it were, to their total contractualisation. I claim that the problem, during this transitional phase, is that our meaningful social relations come to be erroneously regarded as disposable,

an error that brings all manner of ethical complications in its wake. The broader point will be that some of these consequences exemplify the politics of tragedy.

The disposability of contractualised relations is a result of the shift in perceived value just described. The ethical complications have to do with how this shift affects our moral-psychological practices in the wake of interpersonal failure. To get this in view, we need, first of all, to discuss what goes on when parties to non-instrumentally valuable social relations fail to live up to the obligations that constitute them. I argued in Chapter Three that it is a general feature of our reasons for action that our failure to satisfy them makes us apt to feel agent-regret. The thought, roughly, was that if we do not feel agent-regret, it is as if the value of our lives and its constituents is null, making our pattern of feeling inconsistent with our having reasons for action in the first place. The same holds, *a fortiori*, when we breach the constitutive obligations of non-instrumentally valuable social relations. But these relations have an unusual feature: they constitute the value of not only our own lives, but the other party too. Thus, the other party has something at stake in any suggestion of ours that our relationship with them does not in fact matter. For in so doing, we cast doubt on the value of the relationship—a presumed constituent of their life as much as ours. We call into question the meaningfulness of a large swathe of their life, and with it, the quality of that life as a whole. As such, they have a stake in our expressing agent-regret for the transgression, since its absence would suggest that the relationship is a matter of indifference. With such relations, there is something worth salvaging in the wake of failure. Feeling agent-regret befits the value of the relationship, and its interpersonal expression is a first step towards due reconciliation.

Things are different with non-performance in the context of contractual relations. I distinguish two sorts of case. Either the non-performance was on purpose, or it was not. If on purpose, that will be because it has become efficient to do so: owing to situational changes, doing as contractually agreed no longer best serves the interests of the non-performer. But even if it was not on purpose, the fact of non-performance will amount to a situational change for the other party, in light of which it is no longer in their best interests to uphold their end of the bargain. Either way, in the wake of non-performance, the instrumental value of the relationship will be void and the parties will cease to have the performance obligations that once constituted it. So, for the non-performer, there is no obligation hanging around to make them apt to feel agent-regret, and there will be nothing at stake for the other party in the absence of its expression. There is no

relationship left that stands in need of reconciliation; no constituent of the value of the lives of the parties that they need be especially bothered to salvage. The significance of their contract was more fleeting, more fickle, than that. It was nothing more than an instrumental relationship.

Suppose, now, that one party lets the other down in the context of a social relation of some non-instrumentally valuable type. We said that it would be commensurate with the value of that relationship for them to feel agent-regret for their failure. The other party, we said, now has a stake in some expression of agent-regret, in whatever form it needs to take for reconciliation to proceed. Without it, the injury is only compounded. The injured party would be let down twice over: once by the original failure, and again by a subsequent failure to register the original failure by the making of amends. But of course—and this is the point—under conditions of contractualisation, the injurer will see nothing left over worth salvaging: no need for reconciliation, no reason to feel agent-regret, let alone express it. Any obligation that might have made sense of such activity is regarded as having its source in a contract, making its grip on the injurer as fickle as the instrumental value of relationships of that type. As they see it, though their bond may once have been a useful tool, that usefulness has now expired. They see nothing left to do but to dispose of the relationship, which had anyway never brought the parties any closer than strangers to one another, in pursuit of mutual benefit. Contractualisation, I claim, encourages this mistake, exposing already injured parties to the compound injury that attends their treatment as disposable.

These ills of contractualisation exemplify the political dimensions of tragedy to the extent that the constituent obligations of social relations are ‘strict.’ An obligation is strict when what it takes for us to breach them is indifferent to whether we were at fault.⁵ It is the strictness of an obligation that exposes those to whom they apply to tragic situations, in which they may fail, without fault, to meet those obligations, making it fitting to feel agent-regret in response. If those obligations should become contractualised, they will

⁵ That is, strict obligations are those that are not, as Gardner says, ‘fault-anticipating.’ In his words, “Being at fault sometimes contributes constitutively, and not just instrumentally, to the wrongfulness of one’s actions. Some wrongs [...] are ‘fault-anticipating.’ They are wrongs that are committed only in the absence of justification or excuse—in other words, only in the presence of fault” (2005: 115). We follow Stephen Smith (2014: 190) in using the term ‘strict’ in this way. See Chapter Seven for more on this distinction.

be no less strict for it.⁶ But despite the strictness of the resultant obligations, they will still be regarded as having their source in a contract, and will therefore exhibit the fickle nature that characterises the instrumental value of such relations. Notwithstanding their strictness, then, they will be regarded as having lapsed in the wake of faultless failure, so that reconciliatory expressions of agent-regret will not be fitting. There is no possibility of tragedy in contractualised social relations, at least—and however erroneously—as far as injurers are concerned. On the flipside, this leaves those injured in the wake of tragedy exposed to abandonment by their injurer, and needing to pick up the pieces themselves. In this way, I claim, contractualisation shifts the burdens of tragedy in the context of our social relations away from obligors and towards their obligees. I believe this marks a site of the politics of tragedy.

Who, in the social order, can we say that the burdens of tragedy are shifted onto by the creep of contractualisation? The shift, I suggested, is from the obligors in our social relations to the obligees. We may imagine that we can abstract, from the class of obligees, some sensible social grouping with which to do political analysis. I do not think this is so promising. The problem is that, in the contractualised social relations, both parties play a role as both obligors and obligees. To bring this out, I want to consider an example of contractualisation as it affects a particular guise of the social relation between citizens and the state, namely, the relationship between the recipients of socialised care and those who care for them.

I draw on Lydia Hayes's (2017) socio-legal study of the changing landscape of care work in the UK. Many of these changes were subsequent to, and crystallised in, the Care Act 2014, by which "Extensive new duties require local authorities to promote service-user choice and control over care, and also require them to support local care markets" (2017: 159). The result, in Hayes's analysis, was an "exhaustive generalisation of market relations [...] and enterprise in relation to social care," which has "extended beyond the boundaries of social institutions such as local authorities, care homes and hospitals to encompass the individual lives of paid caregivers and people in need of paid care" (2017:

⁶ Or at least not assuming perfect contractualisation as regards the content of obligation. In law at least, performance obligations are strict by default, although it is always possible to effectively build sensitivity to fault into the terms of a contract. Contractualising in this way where the original obligation is strict will amount to what Gardner (2018b: 37) calls the "garbling" of their content. For more on the strictness of performance obligations, see subsection 6.3.1.

158). She speaks of this in no uncertain terms as “a neoliberal epoch of enterprise in the provision of adult social care” which has become “a driver of changed relations between citizens and the state” (ibid.). What were once more durable relations have been contractualised away into arrangements between customer-cum-employers and their paid caregivers, typically women, who “combine informal labour in private homes, [...] work assignments as personal assistants or care workers, and jobs with care companies” (2017: 157).

However vulnerable, recipients of social care owe it to their carers to recognise their labour of care, so long as that labour is undertaken in good faith. So when, under current conditions, that good faith labour is undertaken by a contracted caregiver paid for by the recipient with money distributed to them by local authorities, it is plausible that this recognition is not extended unless the recipient remunerates the caregiver at whatever rate they agreed. Failure to do so is a failure of obligation, one that, again plausibly, will count as such whether or not the recipient is at fault. In that event, it would be fitting for the recipient to feel agent-regret for their failure, and to express it to the contracted caregiver in a reconciliatory mood. But this was not so for one caregiver in Hayes’s study, who took on caregiving work for a client couple. She recounts (2017: 155),

The couple wanted to go on holiday together and I wanted them to have that dream so I took both of them away. It turned out to be an absolute nightmare and I would never do it again. When I got back from the holiday they refused to pay me! I queried it and said ‘You guys agreed to make up the money I was losing from cancelling clients so I could care for you for two weeks solid.’ They said, ‘Well we can’t afford it,’ so I went to the independent living charity that managed their direct payments.

We can surmise from the story that, following the clients’ non-payment, they made no attempt at reconciliation. For them, the relationship did not merit it. It is clear that they regarded their obligation to recognise their carer’s good faith labour as having its source, not in any meaningful relation to their carer, but in the contract. If it was really not the clients’ fault that they could not afford to pay, then that is a tragic situation. But that is a burden of feeling, and perhaps material compensation, that contractualisation has relieved them of. In human terms, this leaves the caregiver doubly injured: not only

does her labour of care go unrecognised; she is subsequently promptly discarded.

This may give the impression that it is caregivers who in general stand to lose out in the contractualisation of their relations with those for whom they care. But that would be too quick. For example, another caregiver in Hayes's study describes one arrangement with a client thus (2017: 156-7).

I started working as a direct employee for a woman who had me for overnights; at least I got money for holidays. Anyway, it didn't last very long because I had to pack in the job when it got so she could no longer move herself. With the equipment and space available I just found it too difficult. When I started, I could pick her up if she fell, but it got to a point where she was too weak and I could not lift her up any longer. I said I would have to stop; I made that judgment.

Primary amongst what is owed by carers to care recipients is an obligation to meet their reasonable care needs, including needs such as the above, that is, of assistance in the event of a fall. Should this become too difficult, what is required is an endeavour to get reasonable accommodations put in place to enable those needs to be met. This is plausibly an obligation that can fail to be met even if the caregiver is not at fault, which is, again, plausibly the case in the scenario described. Though it would be fitting for the caregiver to take steps to reconcile the relationship by an expression of agent-regret, it is clear from the above that no such steps were forthcoming. The obligation to meet the client's reasonable needs has come to be regarded as having its source in the contract, which, following faultless non-performance, has no grip on the caregiver any longer. In this case, as far as the caregiver is concerned, contractualisation has relieved them of the burdens of feeling and action that would be befit their relationship. This time it is the care recipient who is left doubly injured: not only are their reasonable needs left unmet by their caregiver; the caregiver subsequently discards them.

There are, of course, legal protections operating in the background that may, even in the contractualised regime, work to prevent the relief of obligors on either side of the relationship from certain burdens of action. But effective legal protections in this area are, it would seem, more available to those who receive care. It is no secret that there is a crisis of social care in the UK, and Hayes points out that "a steady stream of data about the inadequacy of the social care system in Britain has developed into a torrent" (2017:

28). Properly implemented, this may make welcome the new, careworker specific offence of ‘ill-treatment and wilful neglect’ introduced in section 20 of the Criminal Justice and Courts Act 2015.⁷ Conversely, in the event of non-payment, careworkers only have legal recourse either to HM Revenue and Customs, “where pursuing a formal complaint would not only involve a lengthy wait [...] but would also put the person to whom she gives care at risk of being burdened with heavy fines, tax and wage liabilities,” or to the employment tribunal system where “she would be required to pay a fee,” which at the time of the study would be “typically £390 for minimum wage claims” (2017: 181). In legal terms, then, the contractualisation of these social relations seems, if anything, to favour obligors who are care recipients over those whose care they receive.

But the shifting burdens of tragedy I have been concerned with throughout are not the legal ones, but, as we could say, the human ones. I hope it is clear from the examples that, in human terms, *both* parties are the losers of contractualisation. Generalising to all possible social relations: we *all* stand to lose from their contractualisation. When we let down our important others, faultlessly or not, we see no reason to feel badly about it, no relationship to fix by expressing those feelings. As far as we are concerned, our relationships cannot expose us to tragedy. When we are let down by important others, faultlessly or not, we will find ourselves disposable, with no recognition of the failure or an attempt to make things right. We are abandoned in the wake of tragedy, left to pick up the pieces on our own. These are some of the human costs of the social process of contractualisation. They are testament to how tragedy has a politics; one it is incumbent on us, I claim, as *humans*, to weigh in on with a collective sigh of discontent.

6.3. Against Inflationism in Contract Theory

The previous section sought to contribute positively to the chapter, by offering analysis of the human costs of contractualisation, which, if correct, shows how the phenomenon

⁷ Although, as Hayes brings out, its present implementation exposes poorly-paid caregivers to criminal prosecution, a group in which working-class women, who already face unique pressures in this line of work (2017: 124-5), are disproportionately represented (2017: 198-9). It is a further failing of the current system that legal protections for care recipients cannot be had without further entrenching the socio-economic subjugation of paid caregivers.

exemplifies the political dimensions of tragedy. Subsection 6.2.1 gave an instrumentalist account of the value of our contractual relations, and subsection 6.2.2 went on to give a moral-philosophical analysis of contractualisation based on that account.

I intend for that analysis to stand up on its own terms. But it is important for me to acknowledge that many contract theorists will want to resist the analysis on the grounds that instrumentalism about the value of contract is false. They believe our contractual relationships are, in addition to whatever instrumental value they may happen to have, also of non-instrumental value to our lives, to be engaged with for their own sake. I call these theorists *inflationists* about contract because they seek to puff up the value of contractual relations in our lives, at least relative to my preferred *deflationist* view that the contribution they make is ‘merely’ instrumental.

In this section, I will try to bolster the analysis by a negative route, arguing against contractual inflationism in its many guises. The inflationist tendency typically expresses liberal political commitments in the broadest sense. Such theories fall into two groups. The bulk of them are helpfully regarded as so many versions of what historians of ideas call the *doux commerce* thesis, the Enlightenment view that the commercial way of life—and the contractual relationship that is its foundation—is a civilising force in society, advancing the cause of freedom, pluralism, trust, and other such optimistic values. I will address these theories in subsection 6.3.1. The remaining inflationist theories can be said to valorise the mores of *personal detachment* that arise at a later stage of economic development, expressing a perhaps more pessimistic take on the contemporary liberal polity. I address these theories in subsection 6.3.2. Unsurprisingly, the upshot is that no inflationist theory looks especially promising. I tentatively conclude in section 6.4 that we should after all accept deflationism about contract, and with it, therefore, the analysis of contractualisation offered in the positive part of the chapter.

Before getting started, a quick point on the methodology I will be using throughout the section. I defined as inflationist any theory of contract that says that the contractual relationship is of non-instrumental value in our lives. But not every contract theorist appeals to non-instrumental value thereby advances an inflationist theory, even if their rhetoric has an inflationary feel to it. For their theories to count as genuinely inflationist, it must be that they understand contracts to promote their preferred non-instrumental value *constitutively* rather than merely contingently. To see why, consider the humble cheese grater. We can agree that the cheese grater contingently promotes all manner of

non-instrumentally valuable effects, both by design and by accident, most obviously: the existence of culinary delights and experiences. Presumably this does not imply or even tend to show that cheese graters *themselves* are of non-instrumental value, something capable of directly benefitting us by engaging with them for their own sake. By the same token, a theory that says contracts promote some or other non-instrumental value only contingently does not thereby count as genuinely inflationist. To count as genuinely inflationist, contracts would have to be said to be an essential constituent of the relevant noninstrumental value; something without which the value could not be realised. Many who advance theories of contract with seemingly inflationist intentions would appear, at least sometimes, to miss this value-theoretic point.⁸ There is lots of loose talk of how contract ‘promotes,’ ‘fosters,’ ‘enables’ or ‘enhances’ some or other non-instrumentally valuable sounding thing which, for all its solemnity, remains in the end equivocal as to whether these words are meant to express contingent or constitutive relations. The arguments of this section thus proceed on the assumption that a theory of contract is only genuinely inflationist if the value it appeals to: (a) is noninstrumental, and (b) cannot be realised in the absence of contract.

6.3.1. The *Doux Commerce* Thesis

The French phrase *doux commerce* picks out a strand of Enlightenment thought whose most influential exponent seems to have been Montesquieu, and which was later taken

⁸ I adopt this value-theoretic point wholesale from the aesthetics literature. There, it appears as an objection to ‘empiricism’ about aesthetic value, the view that the value of artworks is explained by the value of our experiences of them. The objection is that, if empiricism is true, then something other than the artwork could have the same value, so long as it stimulates an experience with the same value as the artwork stimulates, a claim that all parties—empiricists included—regard as implausible. They regard it as implausible because it would explain the value of the artwork as instrumental: the artwork would be just one amongst many possible ways of getting the same valuable experience. In suggesting that what makes the artwork valuable is separable from the artwork in this way, empiricists are said to commit what James Shelley calls the “heresy of the separable value” (2010: 710). We could equally say that would-be inflationists who miss this value-theoretic point commit the equivalent heresy as it arises in contract theory.

up by Scottish Enlightenment heavyweights such as Hume and Smith.⁹ In English, it is usually rendered as ‘gentle’ or ‘soft’ commerce.¹⁰ The gist of the *doux commerce* thesis is that the spread of commerce is a civilising force within society, or, more schematically, that the contractual relations that constitute commerce act to promote various benefits of a specifically political character.¹¹ Exactly *which* political benefits commerce is said to act to promote, and exactly *how*, is the messy issue to which the discussion of subsection 6.3.1 addresses itself.

As a preliminary, let me say something to dispel the worry that the *doux commerce* thesis cannot be taken in intellectual good faith, which would, if correct, render serious discussion of it somewhat pointless. Given that the *doux commerce* thesis had its heyday during the apex of the transatlantic slave trade, we could be forgiven for thinking it a fig leaf for European expansionism at best—and a profound artefact of motivated reasoning at worst—which ought not to be taken at face value.¹² I agree it would be an oversight if this thought were not to cross the mind. Nevertheless, when we see the *doux commerce* thesis against its proper historical backdrop it becomes harder to sustain this skepticism. In ideological terms, the *doux commerce* thesis was born of a progressive impulse, and represented, at the time, a push back against centuries of misery and instability wrought by the passions and mores of the *ancien régime*. The promise of *doux commerce* was that, whilst only the aristocratic class get to seek glory and honour amongst rivals, everyone can in principle find betterment in the quiet commercial pursuit of economic interest. The proponents of *doux commerce* rightly saw at their historical juncture the need for and value in an alternative model for the interaction of strangers: one less violent, more popular, and ultimately more amenable to government.¹³

⁹ My presentation of the *doux commerce* thesis leans almost entirely on the authority of the celebrated work of political economist Albert O. Hirschman. For claims about Montesquieu see Hirschman (1997: 60; 70-81), for Smith (1997: 100-13), and for Hume (1997: 66).

¹⁰ This rendering is inevitably less than perfect and is widely regarded as such. For the difficulties of translating ‘doux,’ see Hirschman (1997: 59-60; 61-62).

¹¹ For the claim that the *doux commerce* theorists had in mind specifically political benefits of commerce, see Hirschman (1986: 41-4).

¹² This is perhaps what Marx had in mind when, as Hirschman says, he “recounts some of the more violent episodes in the history of European commercial expansion and then exclaims sarcastically: ‘Das ist der *doux commerce*!’” (1997: 62aa).

¹³ On this inclination of the *doux commerce* theorists, see Hirschman (1997: 63; 69-70).

So, there is a genuine ideological complexity to the *doux commerce* thesis that makes it anachronistic to dismiss as a bad faith proposal unworthy of serious consideration. But the main difficulty with it, I think, is not only its awkward position in the history of ideas. It is, more importantly, tricky to pin down exactly what the thesis is meant to be. The difficulty here is that there may well never have existed anything so univocal as ‘the’ thesis of *doux commerce*, even at the peak of its influence.¹⁴ But for our purposes, there do seem to be a finite number of reasonable ways that we can press the *doux commerce* thesis into something with the shape of a genuine inflationism about contract. My strategy in this section is to consider them one by one. If I can show that they are not individually promising, we can assume that no combination of them will be either.

I believe that the minimum required for an interpretation of *doux commerce* to count as reasonable is that it fits the following schematism.¹⁵ The gist of *doux commerce* is that the contractual underpinnings of commerce act to promote various political benefits with an optimistic flavour characteristic of Enlightenment liberalism. These benefits are macrolevel impersonal values, which range from the less intuitively non-instrumentally valuable (e.g. security, order, peace, wealth, prosperity) to the more intuitively so (e.g. pluralism, freedom, trust). The mechanism by which contract and commerce are said to attain these benefits is via the promotion of microlevel personal values throughout the populace. There are basically two candidates here: some special form of social relations, or some special sort of psychological disposition. Again, the microlevel personal values that are proposed to do the work of *doux commerce* can be more or less intuitively non-instrumental. We will consider plenty of examples of both sorts of microlevel values in what follows. For now, the key point to note is that the values promoted by contractual relations according to the *doux commerce* thesis may lie at either or both the macro- and microlevel. *Doux commerce* may therefore find expression in inflationist theories with quite different shapes, depending on whether the relevant non-instrumental values to be promoted lie at the macro- or microlevel.

¹⁴ As Hirschman puts the point, “What was the concrete meaning of all this *douceur*, polish, gentleness, and even cordiality? Through what precise mechanisms was expanding commerce going to have such happy effects? The eighteenth-century literature is not very communicative in this regard, perhaps because it seemed all too obvious to contemporaries” (1986: 108).

¹⁵ This is intended to approximate Hirschman’s picture (1986: 108-9).

As per the methodological point introduced above, I am only interested in genuinely inflationist interpretations of *doux commerce*, which are those claiming that contractual relations promote—by being constitutive of—non-instrumental benefits of a relevantly optimistic flavour. My discussion will be organised according to *which* noninstrumental value is said to be promoted. Subsection 6.3.1.A considers interpretations that involve the promotion of macrolevel values. Subsection 6.3.1.B moves on to consider microlevel interpretations, focussing on the large class of views that emphasise the promotion of a distinctive kind of social relations. Subsection 6.3.1.C closes by discussing the remaining microlevel interpretation, which emphasises the promotion of distinctive psychological dispositions amongst the populace. I aim to show that none of these interpretations of *doux commerce* amount to a successful inflationist theory of contract.

6.3.1.A. Macrolevel Interpretations

I start by considering interpretations of *doux commerce* according to which our standing in contractual relations acts to promote various macrolevel non-instrumental values. These range from the thinner and less intuitively non-instrumental values recognised by liberals in their more politically realist moments, all the way to the thicker and more intuitively non-instrumental ones that tend to be front and centre in liberal rhetoric. I discuss attempts to construct inflationist contract theories in the spirit of *doux commerce* around macrolevel values of both kinds. To anticipate, the fundamental problem is that the relevant macrolevel values are often not obviously non-instrumental, and that, even where they are, our contractual relations are not plausibly constitutive of them.

Take the thinner and more realist-sounding values like *security*, *order* and *peace*.¹⁶ There is no doubt that security, order and peace in a society is a valuable thing. After all, these are conditions without which it generally becomes hard for people to do whatever they need to do to lead flourishing lives. But explained in this way, these values are not obviously not instrumental. It may be hard for us to flourish without security, order and peace, but if we can do so better by some other means, then, for all the *doux commerce* theorist has said, we lose nothing if we were to make do without them. To this, the *doux*

¹⁶ For the dimension of *doux commerce* relating to values of this kind, see the very helpful paper by Mark Movsesian (2018: 456-7).

commerce theorist can just insist that these values are genuinely non-instrumental. On the simplest way of doing so, we are asked to imagine two equally flourishing societies—one secure, orderly and peaceful; the other volatile, disorderly and in strife—and to concede that the former situation is better than the latter. I struggle, however, to form a clear idea of what I am being asked to make a comparative value judgment about here. What does it look like for somewhere so chaotic to be doing just as well, chaos aside, as somewhere without it? The problem is that the criteria for our concept of flourishing to apply to a society seem to include that it meets a minimal threshold of security, order and peace. This observation invites a more sophisticated line for the *doux commerce* theorist to take. The idea would be that we cannot truly flourish in conditions of total chaos, so that the values of security, order and peace receive a non-instrumental status by virtue of being amongst the constituents of any possible human flourishing.¹⁷ This is to my mind more persuasive as an elucidation of the non-instrumental character of the relevant macrolevel values.

For the sake of argument, let us grant that security, order and peace can be made plausibly non-instrumental by this route. There still remains a fundamental problem for the *doux commerce* theorist insofar as they presume to advance an inflationist theory of contract. That is because, even if it is plausible that the relevant macrolevel values are non-instrumental, it is not plausible that we could not have them without contractual or commercial relations. To think otherwise would be completely blind to the variety of scholarly explanations for the unprecedented levels of relative security, order and peace we are experiencing today, which is sometimes referred to in the international relations literature as the New Peace.¹⁸ Besides scholarly explanations that emphasise nuclear deterrence, peacekeeping institutions, democracy, human and civil rights movements, demographics, and assorted cultural factors, it appears that the dominant consensus, at least in policy circles, is that what fundamentally explains the New Peace is the political

¹⁷ See Julia Annas (1995: 40) for the arguably Aristotelian provenance of this idea.

¹⁸ Though a staple of international relations theory, the term originates in a popular book by the psychologist Steven Pinker (2011). In light of widespread coverage by Western media outlets of the ongoing Russia-Ukraine and Israel-Palestine conflicts, the currency of the New Peace may be thought to be waning. But as far as I am aware, international relations theorists have yet to revisit, on the basis of new data, the justifiability of the continued use of the concept today. I therefore proceed on the basis that their expertise still points to the existence of a New Peace.

hegemony of the US.¹⁹ I am not, of course, qualified to adjudicate between these various explanations. My point is simply that it seems vanishingly unlikely that the sole factor without which we could not have the New Peace is, for all the diversity of scholarly explanations, in the end just global commerce. As such, the New Peace stands as a counterexample to this version of the *doux commerce* thesis.

Before moving on, there is one last thinner and more realist-sounding macrolevel value that I think merits separate discussion. On this proposal, our contractual relations constitutively promote *wealth*, or, if it makes the value sound more plausibly non-instrumental: *prosperity*.²⁰ It is a truism that personal wealth accumulation is merely a means to an end that ought not to be elevated to the status of something to be sought for its own sake. This is less obviously the case at the societal level, in large part because when we imagine a prosperous society, and not just a mass of individually wealthy folk, we likely imagine that the social wealth lifts the lives of all its members in some non-instrumentally valuable way. For the sake of argument, then, let us grant that the value of social prosperity is non-instrumental. The question is whether such social prosperity can plausibly be said to be constituted by contract and commerce, such that it cannot exist without it.

According to the economic historians, the answer seems to be: quite possibly not. Consider Karl Polanyi's (1957a: 18-9) discussion of the Old Babylonian businessmen known as the *karum*. The *karum* acted as middlemen for the mutually beneficial trade of goods between third parties. They did not, however, earn a living through the buying and selling of goods at market prices set by the mechanism of supply and demand. This is because there was no market, and so no prices as determined by the forces of supply and demand within them. Instead, equivalences between quantities of goods the transfer of which the *karum* facilitated were fixed on a long-term basis by decree of a central political authority. The *karum* earned a living from commission taken for their services as intermediaries to third-party trade. This *administrative* trade generated prosperity in the region by stimulating production of goods to be traded at the centrally authorised

¹⁹ For this view of the consensus amongst policy elites, and for dissent from that consensus, see Christopher Fettweis (2017). An extensive review of the international relations literature on the various competitor explanations listed here can be found in Fettweis (2017: 428-431).

²⁰ This dimension of *doux commerce* thought is characterised by Movsesian (2018: 456) as a preoccupation of more recent times.

‘price,’ which they did, in effect, by having institutional status as guarantors of the smooth and timely delivery of the goods desired in return.²¹ In other words, trade in Old Babylonia seems to have taken a predominantly non-contractual form, in that, since the ‘prices’ of goods were centrally determined, there was no place for negotiations and bargaining. The point is that there exists serious scholarship in economic history to the effect that contracts and commerce were not fundamental to the prosperity of the Old Babylonian Empire. To that extent, philosophers who advance the version of *doux commerce* in question can be accused of a sort of disciplinary immodesty, presuming to settle, as they do, empirical debates properly left to the economic historians.

Be that as it may, it might be thought that there is a specific kind of prosperity that really cannot be had otherwise than through markets. Dan Moller says the fundamental benefit of a well-behaved market is that each trade can create value for the parties to it at no extra cost to them (bar transaction costs) besides the lowest possible opportunity cost given their preferences, and can in that sense “create value for free, from nothing” (2019: 122). Now, it is true that certain trades made possible by well-behaved markets—those found along what microeconomists call the ‘contract curve’—non-accidentally create value for free, in Moller’s sense, in a way the administrative trades of the *karum* do not. I do not want to deny that markets really may be constitutive of prosperity in this *ex nihilo* sense. But I do deny that prosperity so understood is non-instrumentally valuable. To see this, consider that talk about how contracts can ‘create’ value ‘for free’ is, in the end, rhetorical, and thus not to be taken literally. All it expresses is that optimal contracts let parties to them satisfy the reasons they each have based on their respective best interests, thus ‘creating’ value, and which reflect their best laid plans for advancing their respective lives, thus creating it without undue opportunity costs, i.e. ‘for free.’ In other words, it expresses that optimal contracts fulfil their point in an optimal fashion, which, as I argued in subsection 6.2.1, is an instrumental one.

I now move onto attempts to give inflationist expression to *doux commerce* in terms of macrolevel values of a thicker and more characteristically liberal variety, like *pluralism*, *freedom* and *trust*.²² Take pluralism first. I do not doubt that toleration and recognition of a diversity of ways of life amongst a populace is a non-instrumentally valuable thing.

²¹ For the language of ‘administrative’ versus ‘market’ trade, see Polanyi (1957b: 262-3).

²² These dimensions of *doux commerce* are emphasised by Movsesian at (2018: 457-9).

But it is, I think, intuitively rather implausible that it is constitutive of a pluralistic society that it be built upon commerce, so that there can be no pluralism in society without its members standing in a system of contractual relations. We can, for example, imagine a large family of people who all get along just fine despite having diverse ways of life (e.g. different religions, values, lifestyles, *Weltanschauungen*), without ever having entered into anything like contractual relations with one another. It will not do to simply protest that their toleration of one another is likely sustained by their familial bonds, since there is, as yet, no reason given as to why a society cannot be relevantly akin to a large family. But suppose, not implausibly, that this cannot be the case in general, since the members of sufficiently large societies will inevitably be for the most part strangers to one another. Do strangers with very different ways of life not need a substructure of contractual interdependence to incentivise mutual tolerance?

Once again, I will look to ancient Mesopotamia for counterexamples, this time from the Neo-Assyrian Empire.²³ According to the royal ideology of the Neo-Assyrian kings, their imperialist expansionism was legitimated by divine command of Ashur, the head deity of the Assyrian pantheon. In practice, however, there is strong historical evidence to suggest that upon imperial conquest of a foreign nation, the Assyrian kings did not force religious conversion upon the defeated. Rather, for reasons of political expedience, they tended not to interfere with local religions, cultic practices and ways of life. In other words, toleration and religious pluralism in the Neo-Assyrian Empire seems to have been conditional on the will of a central political authority, as opposed to the contractual interdependence of its subjects. It may be thought that the deep engagement of the Neo-Assyrian state in international commerce prevents us from ruling out that this was the factor necessary for political pluralism, rather than the good grace of the king. But this reply is inadequate. What is most plausibly relevant to the *doux commerce* theory we are considering is private domestic commerce, the prevalence of which in the Neo-Assyrian Empire is, owing to inconclusive data, a subject of controversy amongst Assyriologists.²⁴

²³ The following counterexample is drawn from the study by historian and biblical scholar Morton Cogan, which is summarised at (1974: 111-3). The same point can be found in the study by Assyriologist Steven Holloway, clearly stated at (2002: 177).

²⁴ For more on the controversy, especially as to why the available evidence is inconclusive, see Nicholas Postgate (2007: 206-7). Admittedly, Postgate himself denies that the Neo-Assyrian government had a monopoly on trade, but he emphasises the speculative nature of this claim, and also presents other speculations from the literature to the opposite effect. The continued

Thus, as an empirical matter, there seem to be good reasons to treat the Neo-Assyrian Empire as a counterexample, and no good reasons not to.

Sometimes appeal is made to thicker and more liberal-sounding macrolevel values in the vicinity of *freedom*, which I will treat as inclusive of such values as *autonomy* and *liberty*. Freedom is a nebulous concept, but it can, I think, be chopped down to size for our purposes on the basis of the plausibility of *doux commerce* theories built around it. The natural tendency, perhaps, is to think of freedom as a microlevel value accruing to individual persons, which we exhibit insofar as we decide what to do for ourselves—on some views *well*—and our acting on that basis is not interfered with by external factors.²⁵ It is, however, implausible that individual people cannot have such microlevel freedom without contract and its surrounding institutions. To see this, just consider how easy it is to imagine an unrivalled and all-powerful despot who possesses microlevel freedom to do exactly as he should so please, without need of anything resembling contract. The usual reply at this stage is that the despot *is* in fact unfree but in the deeper sense that his microlevel freedom to do exactly as he pleases is always conditional on his continued might, which is, in the end, a status he enjoys only contingently. We could put this by saying that he does not belong to a *system of freedom* that guarantees his microlevel freedom come what may. I have a great deal of sympathy for more systemic conceptions of freedom in this vein.²⁶

Systemic conceptions of freedom blur the lines between micro and macro, in that individual people cannot enjoy microlevel freedom in this sense without a macrolevel system that guarantees it. But since the existence of the system is the more basic, let us treat freedom in this sense as a macrolevel value. The question is: can there be a system

state of scholarly dissensus on this point is evidenced by the consistent use of scare quotes around the term ‘market’ as a qualifier of ‘trade’ in Salvatore Gaspa (2018).

²⁵ This gloss is intended to capture what Christian List and Laura Valentini (2016) would call ‘non-robust’ views of freedom in both the narrow sense of the absence of certain constraints, as well as in the wider sense of autonomy, i.e. acting according to laws of one’s own making.

²⁶ These would include the neo-Roman conception of freedom as *non-domination* recently advanced in the spirit of Renaissance civic republicanism by thinkers like Philip Pettit (1997) and Quentin Skinner (2010), as well as the Kantian conception of freedom as *independence*, recently advanced in crystalline form by Arthur Ripstein (2009). On the requirement of a system to guarantee non-domination, see Pettit (1997: 67-8), and to guarantee independence, see Ripstein (2009: 30-1). On the affinity between the two traditions, see Ripstein (2009: 42-3).

of freedom without contract and its supporting institutions? When we survey the philosophical work in this area, the honest answer is: quite possibly not. Arguments to this effect are basically Kantian in spirit. The most promising ones are, it must be said, also Kantian in idiom, which I fear leaves me somewhat underqualified to assess them properly.²⁷ Fortunately, however, such assessment will not be necessary for our purposes. This is because, if it is to be plausibly constituted by contract, then the systemic freedom mentioned in the conclusion of Kantian arguments must have a quite specific *object*. It must concern our ability to decide what to do, with whom, and on what terms—and, of course, not to be interfered with when acting on that basis—when *what* we decide to do happens to be something we cannot do without the other party.²⁸ No matter the details of the Kantian arguments, then, the systemic freedom they must establish as constituted by contract is the very same ability as the one whose value I argued in subsection 6.2.1 is primarily instrumental in nature. So whilst I do not deny that the Kantian arguments might in the end go through, I do deny that the macrolevel freedom they thereby show contracts to constitute is plausibly of non-instrumental value.

Last but not least, I will consider what is in the contract literature perhaps the most popular way of taking *doux commerce* in an inflationist direction based on macrolevel values. On this view, the substructure of contractual relations that underpins commerce is constitutive of something like an ideal form or basis of social relations between people, especially in large societies, in which a mass of people who are more or less strangers to one another need to rub along to get by and live well together. The macrolevel value that immediately comes to mind is something like the value of *trust*, or a *trusting society*, the key feature of which is that people both feel and are able to pursue joint endeavours via

²⁷ The gist of the Kantian arguments, in Ripstein's words, would be that since "the exercise of acquired rights is consistent with the freedom of others," it must be that "anything less than fully private rights of property, contract, and status would create a restriction on freedom that was illegitimate because based on something other than freedom" (2009: 62).

²⁸ Note the contrast with Ripstein, who takes our having contractual rights to be constitutive of our freedom in that their absence would be an "arbitrary restriction" on our freedom to engage in "cooperative activities" (2009: 64). The problem with this is that, as far as parties to contracts are concerned, in the vast majority of cases, the point of entering into them is not to engage in cooperative activity with the other party *per se*, but to do what we happen to be unable to do without them. The freedom of cooperation enabled by contracts is primarily of instrumental significance, that is, we would not be so bothered about its absence if we were able to get whatever it gets us by other means.

trust-based relationships. Other words commonly used to express values in this vicinity would include ‘reciprocity’ and ‘empathy,’ but I will proceed in terms of ‘trust’ and its cognates. There are of course more or less subtle differences between the meanings of these terms in their myriad senses, but they are, I think, faithfully regarded as variations on the same theme.

I do not think that the semantic cloud surrounding trust is obviously evocative of impersonal values at the macrolevel. ‘Trust’ is typically used to pick out something that a person or a group *does*, and when its use seems to pick out something more like a mass of impersonal *stuff*, that can usually be taken as a lyrical way of expressing something about what a group of people generally do. So, whatever the relevant value, I assume it can be analysed without loss in terms of the microlevel value of interpersonal attitudes of trust—and the associated form of social relations—as distributed throughout the relevant populace. Variants of *doux commerce* that suggest contracts are constitutive of trust can therefore be unpacked in microlevel terms. As such, they will receive sustained treatment in subsection 6.3.1.B, where I consider the prospects for inflationist theories of contract based on interpretations of *doux commerce* built around the microlevel value of trust-based social relations.

This concludes discussion of inflationist contract theories built around macrolevel interpretations of the *doux commerce* thesis. We looked at theories focussing on a range of macrolevel values, from the thinner and more realist—e.g. *security*, *order*, *peace*, and perhaps *wealth* and *prosperity*—all the way through to the thicker and more distinctively liberal—e.g. *pluralism*, *freedom* and *trust*. I argued that none of these views is especially promising. In a nutshell: the relevant macrolevel values are in many cases not obviously non-instrumental, and even where they are, it remains implausible that our contractual relations are constitutive of them.

6.3.1.B. Social-Relational Microlevel Interpretations

Microlevel *doux commerce* theories emphasise either the value of a distinctive form of social relations, or the value of something more psychological like a disposition, virtue or habit of mind. I consider these views in turn, in subsections 6.3.1.B and C respectively. For those who take the former approach, the idea would be that contracts and commerce promote a distinctive form of social relations fit to secure the kind of cohesion and

cooperation that brings macrolevel political benefits in its wake, and what makes these social relations distinctive is that they are shot through with interpersonal trust. These *trust-based social relations* are said to be non-instrumentally valuable, and are associated with concomitant forms of *trust-based cooperation*.

One view along these lines can be quickly dismissed. There is wide agreement that a deficit of trust in our meaningful special relationships makes them fare worse *qua* whatever kind of relationship they are, which in turn, and without further ado, makes our lives fare worse to that extent. Given that special relationships are plausibly trust-based in this way, it is possible to take them as the non-instrumentally valuable form of social relations around which to construct this type of microlevel *doux commerce* theory. The claim would be that our standing in contractual relations is a constitutive element of our meaningful special relationships, such that we cannot have these relationships without our having entered into contractual relations. Views along these lines are, I take it, deeply implausible.²⁹ People seem to be eminently capable of standing in meaningful relationships with one another without their ever having entered into anything like contractual arrangements. If anything, as we will see, appeals to contractual mechanisms in the context of such relationships tend to destroy what is distinctively valuable about them. Attempts to build inflationist *doux commerce* theories around trust-based social relations in this sense can therefore be put to one side.

Oftentimes, when appeal is made to the value of trust and trust-based relations, it is ultimately with an eye to the value of associated forms of cooperation.³⁰ Cooperation

²⁹ For subscribers to the *promise theory of contract*, there are ways of understanding promises that would have something like this implausible implication. It follows if the non-instrumental value of undertaking promissory obligations and similar such unilateral voluntary obligations is thought to be constitutive of having or furthering our meaningful relationships, as is a possible reading of, for example, Joseph Raz (1982: 928-9) and Richard Holton (1994: 69). I tend to think this view mischaracterises the role of promises in real life. I agree with James Penner's observation that real life promises are "largely untypical between intimates, and essentially deal with human frailties and pathologies which would otherwise work to undermine agreements" (2014: 124). They are not obviously something without which we could not have or further our meaningful relationships. This is not to deny that promises may, as Daniel Markovits says, "be a path into love," though, as he goes on to say, they are "surely not the only path" (2011: 310).

³⁰ For example, political scientist Russell Hardin opens a chapter of his monograph on trust by declaring that "We are concerned with trust and trustworthiness because they enable us to cooperate for mutual benefit" (2002: 173).

between people does not, of course, have to happen in the context of meaningful special relationships. So, perhaps the idea is that contract is constitutive of social relations that are not meaningful in and of themselves, but which are 'trust-based' in that they enable us to do what we happen to be unable to do all by ourselves. The claim is that contract is constitutive of trust-based social relations and cooperation of a merely instrumentally valuable kind. I believe this claim is multiply problematic. We could first of all deny that social relations and cooperation so characterised would have to be trust-based. After all, all that is required for people to cooperate with one another is that their behaviour be reliable, which can be secured through all sorts of legal, institutional and governmental incentives that need not have anything to do with trust.³¹ We could moreover deny that contract could be constitutive of trust-based social relations at all because, as we will see, contract is not a habitat especially conducive to attitudes of interpersonal trust. But more importantly for our purposes, the claim is dialectically problematic. This is because it concedes from the off that the point of the relevant social relations is merely instrumental. So, even if we grant that contract really does constitute the relevant trust-based social relations and cooperation, the resulting interpretation of *doux commerce* would not support a genuinely inflationist theory of contract.

The failure of this view does, however, suggest another route for the *doux commerce* theorist. They can insist that the trust-based social relations and cooperation they have in mind really are of non-instrumental value. To this end, they could explain the relevant social relations such that, on the one hand, they have a *sui generis* form irreducible to garden-variety special relationships, yet on the other, they plausibly make a genuinely non-instrumental contribution to our lives, such that we should not want to lose them even if we were to figure out a way of getting on without them. The claim would then be that contracts are constitutive of these *sui generis* non-instrumentally valuable trust-based social relations. I believe that something like this shape is shared by a large family of inflationist contract theories expressive of microlevel readings of *doux commerce*.

³¹ This is amongst the central contentions of the cross-disciplinary book by Karen Cook, Russell Hardin and Margaret Levi (2005), in which the authors push back against the "many discussions of trust [that] take cooperation to be virtually defining proof of trust," with the alternative view that "*many interactions in which there is successful coordination or cooperation do not actually involve trust*" (2005: 8).

Prominent views in contemporary contract theory belong in this family, and I shall consider some of their idiosyncrasies in what follows. But to my mind, it helps to think of their resemblances as traceable to a common ancestor in the figure of the *friendly stranger* that first enters the legal-theoretic imagination in the work of Lon Fuller (1969: 27-30). A rough caricature of Fuller's idea is that contracts constitute a novel form of social relations that comes into its own under the pressures of modernity, wherein the vicissitudes of urban life make it such that our dealings are for the most part made with people who are strangers to us. The quintessential feature of these social relations is a certain *openness* of normative expectations, as reflected in the possibility of bargaining and negotiations as regards their contents. This makes them inimical both to contexts of intimacy, where expectations are relatively closed off due to patterns of behaviour established in previous encounters, as well as contexts of hostility, where the disclosure of interests that is necessarily involved in good bargaining and negotiations is mutually disadvantageous. Between intimacy and hostility, these novel social relations represent, in Fuller's words, "a stopping place midway that can be described as the habitat of friendly strangers" (1969: 27).

There is a certain homeliness and familiarity to the Fullerian figure of the friendly stranger. It is readily—perhaps eerily—serviceable for the *doux commerce* theorist. The proposal would be simple. Strangers by definition are not intimates, and so contract cannot, on this proposal, receive its value by being a constituent element of the value of special relationships between intimates. But equally, since the friendliness of friendly strangers indicates the absence of enmity and hostilities between them, the proposal cannot be that contract is constitutive of social relations of an actively evil kind either. The thought would rather be that contract constitutes social relations of cooperation at arm's length, in which friendly strangers deal with one another in something like trust and good faith to secure their mutual advantage. I believe that Fuller's recognition of this distinctively modern form of social relation, with its *sui generis* expectational profile, is an important insight.³²

A basic challenge confronting any *doux commerce* theorist who appeals to the social relation of friendly strangerhood is that, given how Fuller sets the relation up, it is not

³² According to Prince Saprai (2019: 88), this Fullerian insight has been neglected by large swathes of subsequent contract law theory.

immediately obvious why we are supposed to think it is of non-instrumental value. To put the point crudely: simply explaining that something is positioned midway between good and evil is not in and of itself a way of explaining it as good. Supposing we can put this challenge to one side as an artefact of unfortunate rhetoric, there still remains the question of why we should think it non-instrumentally valuable to stand to one another as friendly strangers. Of course, the official Fullerian line would refer us back to the quintessence of these social relations, namely, the expectational openness embodied in the posture of parties to contract formation. That much is clear from the tenor of Fuller's (2002: 203) remarks, when he says,

the creation of a complex contractual relationship through explicit negotiations requires a certain attitude of mind and spirit on the part of the participants [...]. Each must seek to understand why the other makes the demands he does even as he strives to resist or qualify those demands [...]. Explicit bargaining involves, then, an uneasy blend of collaboration and resistance.

In other words, the friendliness of friendly strangers lies in their preparedness to be reciprocally sensitive to the other's interests in course of pursuing their own. The social relations that are alleged to be constituted by contract may not be *friendships*, but they are friendly in that they presuppose attitudes of reciprocity and mutual responsiveness, and are, to that extent, relations underpinned by attitudes of something like good faith and interpersonal trust. It is the friendliness of friendly strangerhood that is supposed to make it a form of social relations that matter to us non-instrumentally. On microlevel interpretations of *doux commerce* built around it, friendly strangerhood is taken to be a non-instrumentally valuable social relation based on trust—or at least seemingly trust-like attitudes of reciprocity and mutual responsiveness—which cannot be had without contracts, and the bargaining posture internal to them.

Explained this way, however, I do not find it at all intuitive that the friendliness of friendly strangers, their mutual attitudes of interpersonal 'trust,' is such as to secure the non-instrumental value of the social relation. No matter how warmly disposed we may be to our contractual partners, the primary point of entering into contracts will always be that doing so ultimately serves our own self-interest. However friendly we may be, it is simply not reasonable for us to enter into contracts we do not judge to be in our best

interest. When we adopt the posture of bargaining required for contract formation, any sensitivity that we exhibit to the interests of the other party will always be conditional on subtle judgments to the effect that the resulting deal would still be favourable to us. The moment we see that no amount of sensitivity on our part will get us a favourable deal, our empathy becomes pointless, and the bare conditionality of our friendly façade asserts itself. This makes our relation superficial in a way that seems to limit the extent to which its subsequent loss can really matter to us. I do not deny that good bargaining thus requires empathically feeling our way into the perspective of the other's interests, to understand their preferences, why they have them, how they are ordered, their red lines, and so on. Nor do I deny that contracts presuppose reciprocity and mutual responsiveness on the part of the parties, and that they thus embody a complex interplay of interests, which, all being well, represents a mutually advantageous settlement. But if this is the 'trust' that lies at the heart of friendly strangerhood, I deny that it is plausibly of non-instrumental value. We extend whatever 'trust' there is here for the sole purpose of securing our self-interest, which, of course, puts that behaviour in a paradigmatically instrumental role.

Some remarks on the history of ideas may help sharpen this point. Although Fuller never explicitly says this, it is reasonable to believe, given his frequent and approving allusions, that his friendly stranger finds inspiration in the sociologist Georg Simmel's early reflections on *the stranger* in society, first published in the German language compilation *Soziologie* (1908).³³ In that essay, the paradigmatic exemplar of the social type Simmel refers to as the 'stranger' is the *merchant*, to whom our bonds are said to be characterised by a kind of looseness and fluidity (1950: 403-6). In the same era, Simmel also writes about how the self-interest of the merchant under market competition can "achieve what usually only love can do: the divination of the innermost wishes of the other" (1955: 55), lending the merchant an uncommon sensitivity to the interests of his customer. Any ethicist would point out that if the merchant shows a sensitivity native to love, he clearly does not do so motivated by it. Simmel's cautious ambivalence about this mercantile basis for social relations was, of course, anticipated less cautiously by Kant in the *Groundwork* (Ak 4:397), where he conveys, with his famous shopkeeper example, a

³³ The essay 'The Stranger' appears there as the third section of chapter nine, and is translated into English in Simmel (1950). Fuller (1969) cites Simmel approvingly at three separate points, mentioning the essay on the stranger specifically at (1969: 9)

more pessimistic attitude towards the value of the merchant's reciprocity.³⁴ Seen from this perspective, it is perhaps surprising that Fuller does not show the same ambivalence to his beloved friendly stranger.

I find it no less surprising with subsequent contract theorists who follow in Fuller's footsteps. The case of Daniel Markovits (2004) is perhaps especially peculiar, given the self-avowed Kantian credentials of his view. Markovits understands the value of the form of social relations constituted by contract to satisfy the ideal of moral community contained in Kant's Formula of Humanity, according to which we should treat persons "*always at the same time as an end, never merely as a means*" (Ak 4: 429). He unpacks the Kantian ideal of moral community as follows (2004: 1432-3). Moral community is established between people when their social relations are such that (a) they have ends that are shared (i.e. they are not inconsistent with one another), and (b) their ends do not just happen to be shared (i.e. their ends are consistent because they each give the other authority over them). Markovits then suggests that contracts satisfy these twin criteria (2004: 1461-3). They satisfy (a) because the parties exhibit the pattern of self-consciously interlocking intentions essential to *joint intentional activity*, which includes, specifically, intentions that favour the efficacy of the other's intentions.³⁵ They satisfy (b) because contractual parties also go beyond what is needed for joint intentional activity, in that they intend not only to perform as contracted but to incur a *directed obligation* so to perform, that is, to place their own performance under the counterparty's authority. This is how Markovits explains the value of the social relations constituted by contract as non-instrumental. Since contracts satisfy (a) and (b), parties to them "treat each other *as ends*, which is to say that they cease to be strangers and enter into a moral community together" (2004: 1463).

I accept Markovits's description of how contracts typically satisfy (a) and (b). But I take issue with his claim that the way contracts are said to satisfy (a) and (b) establish the properly 'moral' and thus non-instrumental value of the resultant social relations. I

³⁴ For Simmel's ambivalence about the self-interested other-regard endemic to modern society, note his assessment that its "liabilities [...] in the social balance sheet must [...] be added to [its] immense synthetic force" (1955: 55).

³⁵ The concept of joint intentional activity is a technical concept that Markovits borrows from the theory of action and intention expressed across several papers collected in Michael Bratman (1999). For Markovits's reconstruction of the relevant points in Bratman, see (2004: 1451-6).

have problems related to both (a) and (b). Regarding (a), as with Fuller, the value of the mutual intentional responsiveness involved in contract *qua* joint intentional activity looks to be straightforwardly instrumental. Markovits does, to his credit, acknowledge this sort of complaint, to which he replies by wheeling in the significance of contractual obligations as per criterion (b) (2004: 1458-9). But there is a further problem with his suggestion that the features of contract in virtue of which it satisfies (b) suffice to secure its non-instrumental value, even in the presence of the features relevant to (a). The claim, we said, is that when we give another authority over our ends, at least where those ends are shared, we necessarily treat the other as an end-in-themselves, lending our social relations with them moral value. I think this claim faces the following counterexample. Consider a case where I promise you that I will ϕ where my doing ϕ is in my interest but not yours (though it is in no way inconsistent with your interests), with the intention that, by giving you authority over me in the matter of my doing ϕ , I make it more likely that I will ϕ . Perhaps I do so because I want to ϕ , though I believe the authority of my own will is weak in the matter of my doing ϕ . My promise satisfies (b) in the same way as Markovits suggests contracts do. But the resultant social relation between us is not, I take it, plausibly of non-instrumental value. For it is straightforwardly a case in which I am instrumentalising you and your authority over me to get what I want.³⁶

These considerations are germane to a group of seemingly quite different views that suggest that what is non-instrumentally valuable about contract is that they extend an *invitation* to trust. Charles Fried (2015: 16) appears to have sown the seeds for this way of talking in the recent contract literature, which figures prominently, for example, in the writings of Daniel Friedrich and Nicholas Southwood (2011: 278) and Prince Saprai

³⁶ It will not do to point out that the instrumental value in my making the promise can sit side-by-side with the non-instrumental value of the resulting social relation, which attends any establishment of interpersonal authority regardless of the reasons for which that is done. For this to work, it would have to be clear that there actually *is* non-instrumental value in the social relation that results, the mark of which would be that we rather it exist than that it does not. But in the present case, this remains a matter of perspective. After all, it is not unreasonable to think that, given that I am instrumentalising you, it only makes matters worse that I am doing so by giving you authority over me, i.e. we should prefer it that the resulting social relation does *not* exist.

(ms.).³⁷ The emphasis on ‘invitation’ aspect is perhaps misleading insofar as it carries the suggestion that the relevant value accrues to contracts primarily because they involve getting the counterparty to *believe* that we will fulfil the moral obligation we voluntarily incur. Getting others to form beliefs with such contents cannot be of non-instrumental value, for if it were, we should want to go around trying to do so for its own sake, which is absurd. This may seem resolvable by narrowing down the relevant beliefs to those the causes of which non-accidentally make them true, i.e. those vindicated as *knowledge*. But this does not improve things. For starters, such a view would imply that a party in breach of contract cannot have invited their counterparty to trust them in the first place, which is both descriptively and normatively implausible. It is also no more obviously non-instrumentally valuable to get others to know non-normative propositions *per se* than it is to get them to believe them: for example, knowledge of mundane propositions, such as the number of coins in my rear left pocket. We might say, instead, that the non-instrumental value in the present case may stem from that of *what* is known, namely, that we fulfil our voluntarily incurred moral obligations. Facts of this kind are usually thought to consist in our standing in relations of Kantian respect, or something like it.³⁸ However, the best developed view along these lines is Markovits’s, which, as we have seen, does not convincingly explain the value of such relations as non-instrumental.

One final prominent and recognisably Fullerian contract theory I will consider is that of Seana Shiffrin (2016). Shiffrin agrees that contracts constitute social relations of something like friendly strangerhood, that is, arm’s length social relations that make a non-instrumental contribution to our lives. But her explanation of this contribution emphasises not so much the bargaining posture internal to contract as its doctrine of awarding damages for non-performance on the pattern of *strict liability*, or, as she says, its “remedial indifference to fault” (2016: 354). The strictness of contractual liability makes contractors responsible not just for making a faultless effort to perform as agreed, but for actual success in doing so. Shiffrin’s key claim is that this works to disburden the counterparty of the sorts of self-protective measures that would make sense under a fault-based liability regime, including, for example, monitoring for the faultlessness of

³⁷ It may be considered a stretch to see Fried as belonging to the Fullerian family, although note the recognisably Fullerian imagery in his more recent remarks to the effect that trust between contractual partners makes them “neighbors rather than strangers” (2007: 8).

³⁸ For the Kantian influence on Fried’s notion of (an invitation to) trust, see (2015: 8n2).

the contractor's efforts. In this way, strict liability in contract makes possible relations and cooperation based on what Shiffrin calls *agency-trust*, in which we give each other "breathing space in which to perform" (2016: 370).³⁹ Contract is in this sense said to be constitutive of trust-based social relations and cooperation.

Agency-trusting social relations may well count amongst their contingent effects a variety of noninstrumental benefits. These include, most obviously, noninstrumental benefits arising from a more efficient allocation of resources, i.e. less time and energy wasted on self-protective measures, including time and energy wasted on any conflicts that result from measures that give the other party insufficient room to breathe (2016: 365-6). But for the resulting reading of *doux commerce* to support a genuine inflationism about contract, the relevant social relations would have to be essentially constitutive of these benefits, which seems unlikely given how easily we can cook up alternative ways of getting the same results. This is, I think, why Shiffrin instead brings agency-trust under the value of respect for *privacy*, which she suggests is a "relational consideration" of a kind that can ground a properly "moral justification" (2016: 364-5).⁴⁰

An initial problem with the appeal to privacy is that, at least understood broadly, it is simply not plausible that we cannot respect it without agency-trusting relations. For instance, the relations between complete strangers who do not intend to work together need not be based on agency-trust, but presumably they can and should respect each other's privacy in a broad sense. On closer inspection, however, it appears that Shiffrin has something narrower in mind with 'privacy' which more plausibly cannot be had without agency-trust. She has in mind a constituent of what she calls "morally healthy dynamics in relationships between contractors" (2016: 379), that is, specifically between strangers who actively seek to cooperate in pursuit of mutual advantage. I do think it is more plausible to appeal to privacy in this narrow sense, since it is easy to see, without agency-trust, how cooperative relations between self-interested strangers could become

³⁹ According to Shiffrin (2016: 365), agency-trust is a *sui generis* kind of trust that should be distinguished both from *epistemic trust* (i.e. where the trustee is trusted to represent facts and their intentions accurately) and *fiduciary trust* (i.e. where the trustee is trusted to have the trustor's best interests at heart).

⁴⁰ Sometimes Shiffrin speaks of 'autonomy' in this connection, which she seems to treat as roughly interchangeable with 'privacy'. I focus on privacy, since if what she means by 'autonomy' is something more than this, then it presumably falls foul of considerations given in subsection 6.3.1.A under the heading of 'freedom.'

objectionably stifling, meddlesome, or in a word: *micromanagerial*. That being said, the value of privacy thus conceived is rather unglamorous. It makes for relations between strangers whose ‘moral health’ consists in nothing more than that they have built-in safeguards against the more vicious excesses of self-interested behaviour. I do not deny that such relations may be valuable, but is engaging with them really a direct source of enrichment in our lives, to be pursued for their own sake? I cannot see how. Relations whereby we ought not kill each other (say) are also ‘morally healthy,’ in a sense, but our having and meeting their demands does not make our lives better *per se*. They are only such as to make our lives *pro tanto* worse if we do not.

These considerations dovetail with a more basic objection to Shiffrin’s view that applies insofar as it takes a broadly Fullerian shape. The worry would be that, just as the mutual responsiveness of Fuller’s contractual bargainers is in the end circumscribed by their respective self-interest, so too, for all Shiffrin has said, may be the respect that strictly liable contractors show for each other’s privacy through agency-trust. Once it is conceded that the privacy of agency-trusting relations is not of non-instrumental value, there is little reason to think that agency-trusting relations make a non-instrumental contribution to our lives at all. After all, agency-trust is a corollary of strict liability in contract, and the primary point of entering into contracts is to secure our best interest when we happen to be unable to do so all by ourselves. This suggests, if anything, that we cooperate under the regime of strict liability in contract because and to the extent that an absence of micromanagement is to our mutual advantage given what we are trying to achieve, i.e. that we extend agency-trust based on the outcome of a calculation. That explains why it can sometimes make sense, as Shiffrin acknowledges,⁴¹ for one or both parties to build fault-sensitivity into the contractual terms, as could be the case, for example, if they take the other’s possibly meddling in their affairs to be preferable to being lumped with the costs of a faultless accident. For all these reasons, I claim, agency-trust does not look to be a sound basis for microlevel social relations fit for inflationist contract theories expressive of *doux commerce* so interpreted.

I think that with the failure of the Fullerian views, the prospects for inflationist *doux commerce* theory built around trust-based social relations look rather bleak. But it gets

⁴¹ Practical agreements based on a fault-liability rule, Shiffrin says, would be “morally coherent, and [she is] not suggesting we should obstruct parties from contracting around the strict liability rule” (2016: 369).

worse. Trust-based views of contract in general face the following deep objection, which I have mentioned in passing several times now. The objection says that contracts are by their nature an inhospitable environment for attitudes of interpersonal trust, and so they cannot be constitutive of whatever value there may be in trust-based social relations. According to Dori Kimel, what makes contract inimical to trust is its *enforceability*, which, as he puts it, “casts a thick and all-encompassing veil over the motives and the attitudes” of the parties, leaving the many facets of contractual conduct “largely devoid of expressive content” (2003: 74). This in turn makes contract “a singularly inadequate framework in which to express attitudes and reveal personal attributes of the kind that tend to create and enhance personal relationships” (2003: 77), of which, he says, “trust and respect are the most fundamental building blocks” (2003: 29). In other words, the fact that conduct within the sphere of a contract takes place under the watchful eye of the state makes it impossible for contractors to be certain that they do what they do out of genuine trust for one another, as opposed to fear of state coercion.

There is something intuitively right in the idea that legal enforceability can crowd out respectful motives and attitudes of trust. But equally, we do not ordinarily suppose that common knowledge of the fact that our interactions are governed by legal norms enforceable by a violent state apparatus suffices to justify a generalised suspicion about their motivational basis. For instance, Shiffrin (2012: 253) points out that,

In the case of tort, there is little serious concern that legal regulations on bodily contact have come to dominate the motives of citizens or that citizens believe their safe passage across the streets is generally a matter attributable to law and not to basic civic decency. Why should we worry more about the case of contract law’s infiltrating and tainting moral citizens’ primary motives of promissory fidelity and trustworthiness, rather than its working as a form of official recognition and source of backstop assurance?

Though I see Shiffrin’s point here, her question can, I think, be answered in the spirit of Kimel’s remarks. For any given interaction between citizens, it is governed by tort law

norms whether they like it or not.⁴² Thus the mere fact that an interaction is governed by tort law norms does not thereby give citizens any presumptive reason to think that they prefer their interaction to be so governed, as they might, for example, if they did not antecedently trust each other to interact out of respectful motives. So, even if it is true, as Kimel implies, that the background fact of tort law makes citizens less certain about the motives behind their interactions, it does not supply a presumptive reason to doubt them. Compare: I may literally be uncertain about what is round the corner, but, ordinarily, I have no reason to doubt that it will not be an axe murderer.⁴³ The fact that there is a violent state hovering around therefore need not be inimical to trust. But things are different where it is up to citizens whether or not their interaction is governed by legal norms, as is the case in a contractual context. Where this is so, the mere fact that an interaction is governed by legal norms *does* justify the interactors in presuming that they do not antecedently trust each other to perform as agreed. After all, that is what explains why they decided to invoke the violence of the state when they had the option not to. In other words, though the omnipresence of a violent state need not be inimical to trust-based relations between citizens, it surely is when they actively appeal to that violence to underwrite their interaction.

So, whilst I agree with Kimel in spirit, I believe he slightly misdiagnoses the issue. Interpersonal trust may flourish in the long shadow cast by the state, but not between those who would rather not interact without appealing to its coercive apparatus. Once that appeal has been made, as in contract, I see no problem with Kimel's explanation as to why the resulting environment of incentives is not conducive to the flourishing of trust between them. But I do not think this is the basic explanation for why there is no

⁴² It is, of course, up to would-be tort claimants whether or not to pursue litigation, so that in this narrow sense, it is up to them whether their interaction is 'governed' by tort law norms. But what I have in mind here is a wider sense, in which an interaction is 'governed' by norms when the interactors are in principle entitled to enforcement under them (whether or not they actually get enforced).

⁴³ This is the key point of my divergence from Kimel. He suggests that it is enough for legal enforceability to crowd out trust that, when it comes to motives, "the situation is typically one where we simply do not know" (2003: 76). My point is that there are lots of things we do not know, or of which we cannot be certain, that we nonetheless have no reason to doubt.

trust in contract.⁴⁴ There is no trust in contract because the whole point of appealing to the contractual mechanism in the first place is to allow for cooperation between parties for whom there is an antecedent deficit of trust. That is why the fact of enforceability in contract justifies a presumptive doubt about respectful motives, in a way that does not generalise to tort law settings or indeed to law-governed interactions more generally. Kimel offers a compelling explanation for why the enforceability of law quite generally can make it difficult or even impossible to (re)establish trust, *if and when doubts have set in*. But to explain why contract specifically is inimical to trust by default, we need to add that such doubts are presupposed by contract, in that the contractual mechanism is there as an option to make up for them.

6.3.1.C. *Dispositional Microlevel Interpretations*

Subsections 6.3.1.A and B together show that inflationist contract theory expressive of *doux commerce* cannot be built around macrolevel values or microlevel social relations, because the candidate values are either (a) not plausibly non-instrumental, or (b) not plausibly constituted by contracts and commerce. I will now stop to consider the third and final shape a reasonable such inflationism can take, which appeals to the microlevel value of something more psychological like a *disposition* or *habit of conduct* or *of mind*. Specifically, the claim would be that contract is constitutive of some non-instrumentally valuable such habit, also known as a *virtue*.

Not surprisingly, the habits that make the best candidates are those that are at once plausibly advantageous in a commercial context, whilst also bearing some facilitative connection to macrolevel values of a liberal-optimist bent. An emphasis on these habits, which I will call the *mercantile virtues*, is perhaps most in the spirit of the classical *doux commerce* thesis, since they are precisely those in virtue of which commerce is said to be ‘doux.’ Many of the candidate mercantile virtues such as trust, reciprocity, empathy, and

⁴⁴ This may explain why Kimel’s stock example of parties whose appeal to the contractual mechanism in the matter of their special relationships is an atypical case of contract. I agree with Kimel’s assessment (2003: 56), shared by John Gardner (2018a: 44), that this is usually indicative of a breakdown of trust in those relationships. But my point is to emphasise that the breakdown of trust is *upstream* of the appeal to contract, even if enforceability makes for a continued lack of downstream trust as well.

so on, are in effect psychologised versions of the values of their cognate social relations. For that reason *doux commerce* theories built around them can be rejected on the same sorts of grounds as we considered in section 6.3.1.B. Others of the candidate virtues, like *open-mindedness*, *fair-mindedness* and *orderliness* seem to be psychologised versions of some of the macrolevel values we considered, in particular, pluralism and security, and can be rejected on the grounds considered in section 6.3.1.A. Yet others like *reliability*, *thrift* and *prudence* have so far not appeared, though to the extent that they even strike us as non-instrumentally valuable in the first place, it is clear, I think, that they can be swiftly rejected on the grounds that it is just not plausible that they cannot be inculcated in a populace except through their contractual dealings.⁴⁵

To my mind, a far more promising way of appealing to the mercantile virtues has a quite different structure. These views start off by conceding that the role of contract in our lives is primarily the instrumental one of securing our best interest, when the way to do so happens to be something we cannot do all by ourselves. They then observe, in my eyes correctly, that if contracts are to play this instrumental role then they need to be binding, that is, they must be a source of obligations to perform as agreed. The final move—and this is where the mercantile virtues come in—is to say that the bindingness of contract presupposes a certain psychological milieu, in which people are in the practice of *contractual fidelity*, that is, where feelings of honour are associated with the unflinching fulfilment of bargained agreements typical of a modern, market-based way of life. On views like this, the mercantile virtue of contractual fidelity is part and parcel of the contractual relationship (via its constitutive obligations) and the commercial way of life that surrounds it. The case to be made for there being a virtue of contractual fidelity is thus the same instrumental one, namely, that it allows strangers to cooperate to their mutual advantage. Nonetheless, the thought goes, this does not exclude that it can be a constituent of a non-instrumentally valuable way of life to show contractual fidelity, or not to fail to show it, at least in those cases where the contract plays its proper instrumental role.⁴⁶ Contract is in this way said to walk in lock step with the mercantile virtue of contractual fidelity.

⁴⁵ These are virtues alluded to at various points in Nathan Oman (2016), helpfully collected in Movsesian (2018: 453–4).

⁴⁶ This view of the value of contractual fidelity mirrors Gardner on the value of what he calls *solidarity*, who, unpacking an Aristotelian thought, writes, “True, the value of the use to which

This appeal to mercantile virtue is subtle, and the inflationist *doux commerce* style contract theory that results is not unattractive. Attractive though it may be, however, it is not without its problems. First of all, many would disagree that it is something purely psychological in virtue of which contracts are binding. As we saw in 6.3.1.B, it is plausible that part of the point of contract is to get strangers to do as agreed in the absence of favourable psychological dispositions, and, as such, the bindingness has to be supplied by the external incentive of the threat of state coercion. So, the idea that contracts cannot be binding without a psychological disposition like contractual fidelity seems flawed. Secondly, as with any putative virtue, the mercantile virtue of contractual fidelity is wide open to political critique. I believe it would be perfectly reasonable to suppose that insistence on contractual fidelity, just like insistence on chastity, is nakedly ideological. Sometimes I even think we can see exactly who this ideology serves. During the 2008 subprime mortgage crisis it was ordinary households whose contractual fidelity needlessly cost them life-destroying sums of money, costs that often directly benefitted large financial institutions who, in turn, were happy to breach their own contracts (both to homeowners and to investors) as soon as they judged them to be suboptimal.⁴⁷ I find it hard to see at what exalted altar ordinary people are supposed to have made their sacrifice, or to ignore who was its primary beneficiary.

But even if the *doux commerce* theorist can find a workaround for these objections, there remains, I think, a deeper problem with the resulting contract theory insofar as it has inflationist aspirations. The claim, recall, is that the contractual relationship receives its non-instrumental value from that of the contractual fidelity that is part and parcel of its constitutive performance obligations. If that is so, then we should expect that what makes non-performance immediately bad for our lives is a corresponding deficit of contractual fidelity on our part. The problem is that this explanation fails to capture what goes on in general when we fail to meet the demands of relationships that matter to us non-instrumentally. To see this, consider that, for such relationships, the failure to meet

something is put counts only towards the instrumental value of that thing. But not all value that depends on the use to which something is put is the value *of* that use. There is also the value that cannot but be invoked and relied upon, in the process, by those who put that thing to use” (2002: 505).

⁴⁷ See the lawyer and legal scholar Matthew Seligman (2019: 888-9) for a recent summary of the evidence on this. Some of the figures are pretty astonishing.

its demands—and the subsequent damage done to our relationship—is going to make our lives go worse in ways that get reflected in the phenomenology and its associated behavioural expressions. This was one of the points I laboured in 6.2.2. Amongst these will be the expression of agent-regret, geared towards making amends and fixing the relationship, which is, after all, a source of non-instrumental value in both our lives. Whatever makes our failure to meet its demands bad for us must be such as to make sense of the reconciliatory behaviour we exhibit in its wake. But if what is purported to make breach of contract matter to us non-instrumentally is no more than a deficit of personal virtue, then even if this comes associated with a negative phenomenology, it is not one that obviously finds sensible expression in the making of interpersonal amends.

I take it that this argument applies to any appeal to mercantile virtue insofar as it has this subtle structure, whether it be contractual fidelity or something else. It has to, if there is to be a distinct moral-philosophical role for the concept of ‘virtue’ at all. On the one hand, if it is not to collapse into the concept of a prudential habit, then it must be such that exhibiting it makes our lives go to that extent better without further ado. But on the other hand, exhibiting a virtue must make only *our* lives go better without further ado, even where its content is such that it cannot be exhibited without taking another person as an object, as in, for example, generosity. That is, if there is a virtue of generosity, then it is conceptually possible that it is exhibited, thereby making the giver’s life better, *without* thereby making the recipient’s life better. Perhaps it will be said that the concept of these interpersonal virtues is such that they are not truly exhibited where doing so does not make life better for their objects. But that would, I think, leave our exhibiting them indistinguishable from our simply meeting the demands of special relationships. So, the *doux commerce* theorist cannot reply that exhibiting contractual fidelity includes the making of amends where things have gone awry, i.e. that it is one such interpersonal virtue. For if they do, then either our amend-making does not necessarily have a chance at making things go better for both of us, which is false to the phenomenology, or what they call contractual fidelity is in fact a form of social relations.

6.3.2. Pessimistic Inflationism

Let us take stock. The point of section 6.3 as a whole is to bolster the section 6.2 analysis of contractualisation by arguing against the different guises of inflationism, a popular

and broadly liberal view in contract theory. Subsections 6.3.1.A to C dealt with theories expressive of the *doux commerce* thesis, which share an emphasis on the kind of liberal-optimist values characteristic of Enlightenment thought. We found them to be wanting insofar as they purport to be genuinely inflationist, because the values they emphasise are either: (a) not plausibly non-instrumental, or (b) not plausibly unrealisable in the absence of contractual relations.

In subsection 6.3.2, the final substantial part of the chapter, I consider inflationist contract theories of a quite different sort, built around the rather more pessimistic—but still decidedly liberal—value of *personal detachment*. This form of inflationism reflects a move beyond Enlightenment optimism about the liberal socioeconomic order, and seems, if anything, to represent something like the obverse of the *doux commerce* thesis. Before we begin, it will be worth getting the idea of pessimistic inflationism into clearer view.

I characterised pessimistic inflationism as the ‘obverse’ of *doux commerce*. This way of speaking is not my own. It is how political economist Albert O. Hirschman describes what he calls the “self-destruction thesis,” according to which, far from promoting the genteel good manners and civility necessary for peace and social order, the commercial organisation of society in fact “carries within itself the seed of its own destruction” (1986: 111). Though, as Hirschman makes clear, the self-destruction thesis takes many forms, the basic criticism of *doux commerce* that runs throughout is that it mistakenly appraises the whole historical dynamic of commercialisation on the basis of what are in fact merely the effects of its honeymoon stages (1986: 109-17). So, even if it is true, as *doux commerce* predicts, that the spread of commerce may initially work to erode the old bonds and hierarchies of an ungovernable feudal system, it is also true, inevitably, that the same forces tend to hollow out the social bonds that lie at the foundations of the new market society. In the end, the logic of commercialisation is thus said to leave the atomised denizen of late modernity without even the minimal morality needed to prop up its own way of life. This critique has recognisable echoes of both conservative and Marxian forms of discontent with the liberal socioeconomic order, though, as Hirschman notes, it has been espoused by thinkers whose discontent took neither form (1986: 110).

It may thus come as a surprise that a contract theory that represents the ‘obverse’ of *doux commerce* could also, in its own way, be expressive of values of an undeniably

liberal flavour. To see how this could be, we need only recognise that it is always possible to lean into criticism and to own it. Liberals who adopt this strategy are not afraid to look their opponents in the eye whilst spinning the target of their criticism into a virtue. The idea would be to find something in the commercial dissolution of our social bonds which is, on closer inspection, in fact a good thing to be welcomed for its own sake. To my mind, a liberal contract theory built around some such value would be more robust than any alternative based in *doux commerce*, since it would express a more sober and nuanced take on just what it is about how we have organised ourselves today that we all, no matter our political persuasion, cannot but agree we have benefitted from in some way. Quite possibly it will convey a more pessimistic image of the merits of the late modern liberal-capitalist polity, but it will, in my view, be the sounder a liberal contract theory for it.

The most advanced proponent of this strategy in contemporary contract theory is Kimel (2003). Kimel is concerned, dialectically speaking, to offer a liberal corrective to what he sees as a tendency in liberal contract theory to overegg the continuity between contracts and promises. He pursues this aim through a careful comparison of the two. Amongst their key similarities is the instrumental function of facilitating reliance of a certain kind, which enables us to do things we happen not to be able to do on our own. But it is a similarity that brings at least two crucial differences in its wake (2003: 65-6). The first is the mechanism by which the practices secure their instrumental function. Roughly speaking: whereas promises typically operate via a thick form of trust native to more involved or intimate kinds of relationships, contracts proceed instead via the more impersonal incentives associated with legal enforceability. The second is a difference in the purported non-instrumental significance of the two practices, which is as an artefact of the differences, just mentioned, between how they respectively secure their common instrumental function. Whereas the trust we invite through promises is said to enhance our relationships, the less involved way of going about things in contract embodies the quite different non-instrumental value Kimel calls ‘personal detachment,’ which is, “in a sense, the diametrically opposed value” (2003: 78).

I believe that there is much to admire in Kimel’s work, and it is dense with kernels of thought that repay turning over again and again. I will have to artificially narrow the focus to what he says about personal detachment. Making plausible the non-instrumentality of the value of personal detachment seems to be the key to establishing

the liberal credentials of Kimel's theory at a depth sufficient for it to play the dialectical role he wants it to. I must admit, therefore, to finding it odd that he devotes so few words to this task. Though what little he does say is characteristically replete with possibilities, I do not think, in the end, that there is enough in there to get the job done.

I will discuss various understandings of the value of personal detachment available in what Kimel says, applying to the resultant contract theories the method we have used throughout. Subsection 6.3.2.A starts by laying aside the clear nonstarters. Subsection 6.3.2.B goes on to consider a more promising understanding, according to which the value of personal detachment lies in how having it as an option deepens the meaning of our existing relationships. I will argue that, though contractual relations, on this view, have an emancipatory potential, the view faces insurmountable political challenges. In subsection 6.3.2.C, I offer, in a conciliatory mood, a story about the connection between values and technology that may allow us to salvage the emancipatory attractions of this view. But the story casts personal detachment in the role of a technology, which is of instrumental value, and so the resulting contract theory, I conclude, turns out to be deflationist.

6.3.2.A. Clear Nonstarters

For Kimel, we saw, contracts enable us to do things by agreement with others without having to get personally involved with them. But what exactly does Kimel have in mind? He begins with the following gloss. The value of personal detachment, he says (*ibid.*),

consists in the very framework contracts provide for doing certain things with others not only outside the context of already-existing relationships, but also without a commitment to the future prospect of such relationships, without being required to know much or form opinions about the personal attributes of others, and without having to allow others to know much and form opinions about oneself.

No doubt the tacit sentiment here will be all too familiar to the twenty-first century metropolitans that can safely be assumed to make up the bulk of readers. For my part, after a busy day, I like that I can immediately put Spotify in my ears, buy a soup with a

minimum of niceties, and get off the bus without need of a ‘thanks.’ I like that these transactions do not have to be interactions.⁴⁸ I do not want to deny that such detachment can be a good thing. But when we reflect on what might lead us to want such detachment from others, it becomes clear that the familiar sentiment harbours all the pessimism—some of it quite possibly justified—of how we think of the mass of strangers out there in late modern society. We basically think of them as unlikely to really warrant a personal investment of time, effort and other such limited resources, not least because it would be better allocated to the people and projects we already have in our busy lives. So understood, the value of personal detachment looks plainly instrumental.⁴⁹

To make all of this more concrete, I suggest that we unpack the pessimism of the sentiment and sharpen it as follows. What sorts of generic worries might be intelligible when we are faced with the prospect of sustained interaction with a complete stranger? Here is an indicative list: (a) they might pry or meddle in our affairs, (b) they may form unwanted expectations of us going forwards, (c) we may simply not get along, and (d) the whole exercise will consume time, effort and other such limited resources. The list aims to be exhaustive, and the intention is for each item to be interpreted such that they are ordered in increasing strength, in the sense that each subsequent item entails the preceding one but not vice versa. I submit that for each of the worries on the list, if it really is good for us to be able to be rid of them, then that value will be instrumental.

We can group (a)-(c) together on the basis that they are worries about possibilities, rather than certainties, as in (d). Note, first of all, that what we worry about with (a)-(c) is not always unwelcome in our relations with others. After all, the possibility of these

⁴⁸ That is, after all, what they are. Maybe these preferences will be thought repellent, in that they express a generally rude or impolite attitude towards the people involved. I can only reply that, towards the end of a double shift in the service sector at a not quite living wage, I also liked it when my customer transactions came with a minimum of niceties and small talk, and perhaps even resented the expectation, from the many-hundredth customer that week, that I perform for them what it could not any more be for me: a meaningful interaction.

⁴⁹ It seems to me that we can, with a view to the instrumental benefits just mentioned, interact with others in a personally detached way without that necessarily being mediated by contracts. For example, at least part of the point of interacting via Reddit forums and other similar online spaces, rather than ‘in real life,’ is plausibly—perhaps conditional on our estimation of online communities—to glean the very same instrumental benefits of personal detachment. As such, I do not believe that contract is an essential constituent of the instrumental value of personal detachment, as Kimel may be taken to suggest in the quoted passage.

things happening, at least on occasion, seems to be part and parcel of what makes our most intimate relationships worthwhile: our being friends, family, lovers, and the like. These relationships could not have the meaning for us they do if they left no latitude for the sorts of annoyances that stem ultimately from our caring about the other (who they are, how they live, how they see us), as are captured under (a)-(c). This lends our meddling, holding to account and falling out a special meaning—even if it really might not feel that way at the time.

It seems to me that if we should want rid of these worries at all, we need to interpret them more specifically. Their object cannot be so much that (a)-(c) could occur *per se* as that people who we are not interested in could subject us to these things. Put this way, we would be rid of the worries if we could control for ourselves who it is that we become personally involved with. The thought would be that we have noninstrumental reason to want the relationships best suited for us, and that, since we have privileged access to who we are and what makes us tick, our having control over who we do and do not get to know more intimately is a good way of getting to have the relationships that are best suited for us. I emphatically agree that this sort of control that comes with the option of detachment can be a good thing. But the value, so explained, is surely instrumental, i.e. to having relationships that are better suited to us.

Now let us move onto (d), the worry that our interaction with the complete stranger will take up time, effort, and other scanty resources. Since we humans are finite beings, any interaction whatsoever—strangers, acquaintances and intimates alike—is certain to have this consequence. For this reason, it is harder to accuse someone with worry (d) of the sort of pathological risk-aversion that worries (a)-(c) may express. And yet, as with (a)-(c), it is not that what we worry about under heading (d) is always unwelcome. For it is, again, part and parcel of what makes our more intimate relationships worthwhile not only that we can be expected to find the time, effort and resources needed to keep them vibrant, but that, at least on occasion, we actually relish in the opportunity to do so. In this context, someone who claims to have found the ‘solution’ to the expenditure of time, effort and resources betrays a fundamental confusion as to the noninstrumental significance of their intimate relationships.

So once again, if we should want rid of worry (d), then its object cannot be that we will consume our limited resources *per se*, so much as that we waste them in so doing. In the first instance they would be wasted on people who do not matter to us personally,

or to whom we do not personally matter.⁵⁰ More broadly, and derivatively, our resources would be wasted on people with whom the prospects of a deep relationship are doomed from the start, or, at least, those for whom we judge that to be probable. In other words, in addition to helping us have relationships that are better suited to us, our control over our involvements also helps us to efficiently allocate time, effort and other such scanty resources to the people who really merit their consumption, as adjudged by those best positioned to do so: ourselves. Therein lies the value of personal detachment, our ability to remain at a cool distance from those on whom we think our resources will be wasted. This seems to be yet more instrumental value: it helps us get the most out of the sources of non-instrumental value we already have in our lives.

Kimel's initial gloss therefore does not obviously help his inflationist contention that relations of personal detachment are non-instrumentally valuable. He does however go on to say more about what he has in mind (2003: 79).

Not only is it easy to see that, when examined in isolation, both personal relationships and personal detachment can be, in the right circumstances, a good thing, but in fact when the two co-exist as options in people's lives, their respective values tend to be mutually reinforcing.

There are at least two thoughts in this passage that may be taken to have a bearing on the alleged non-instrumental value of personal detachment. One seems to me to be somewhat confused, whereas the other is, I think, quite profound. Let us first tackle the confused thought, so as to avoid any distractions it may cause.

When Kimel says that both personal relationships and personal detachment can be a good thing, this will, of course, be both trivial and unhelpful to his cause if 'good' is to be read instrumentally. But on a non-instrumental reading there arise problems as soon as we factor in for the suggestion, as Kimel says, that both relationships and detachment are a good thing "when examined in isolation." The very idea of non-instrumental value seems to be such that if something has it, then it would have it if examined in isolation,

⁵⁰ This would include resources wasted as a result of being shafted by people to whom, as it turns out, we do not and perhaps could not really matter, e.g. due to forms of discrimination. For marginalised people to extend their resources (temporal, material, emotional) to strangers in a society that marginalises them will in all likelihood be a waste of them.

that is, independently of the non-instrumental value of anything else.⁵¹ A consequence of this is that even if nothing else has it, so that, in that sense, we are *dependent* on it as the sole source of non-instrumental value in our life, it will not, just because we happen to be dependent on it, cease to be of non-instrumental value. Suppose, for example, that we were to get stranded on a desert island with just one friend, so that we are dependent on that one friend as a source of the value of personal relationships in our life. There is no doubt that our social life on the island would be impoverished, at least as compared to the vast majority back in civilisation. Indeed, it is likely, on some reasonable measure, to count as absolutely impoverished. But our dependence, I take it, does not necessarily abnegate or even tend to diminish the non-instrumental significance of the friendship. As a conceptual matter, it seems that what is of non-instrumental value remains so when examined in isolation, and is insensitive to our dependence on it in that sense.

The confusion becomes apparent when we recall Kimel's claim, in the above, that both personal relationships *and* personal detachment can be of non-instrumental value when examined in isolation. Though he says nothing that obviously contradicts the first conjunct,⁵² he appears, in light of what we have just said, to go back on himself regarding the second. As he says in several places, what is of non-instrumental value is not so much our standing in relations of personal detachment *simpliciter*, but rather our having the *option* of proceeding via such relations. "Detachment," he says, "is valuable as an option, not a predicament" (2003: 79). I think this idea is attractive and can be spelled out as follows. Personal detachment can be a good thing, but if, in our dealings with others, it is not open to us to proceed in a more involved way—i.e. if detachment is not an option but a necessity—then that makes it seem no good. For instance, those who are socially marginalised may have no choice but to invoke the machinery of law because the others with whom they must deal tend to prove untrusting, untrustworthy, or indeed

⁵¹ Or at least apart from there being noninstrumentally valuable valuers around to value it. More details on 'regress' arguments to this effect are helpfully collected in L. Nandi Theunissen (2018), which also offers an unorthodox proposal on this traditional pattern. As far as I can tell, this point does not affect the thrust of the present argument.

⁵² Kimel also says things that appear to support it, such as, for example, that "certain kinds of personal relations do not depend for their value on the parties' unlimited freedom to pursue or to mould them" (2003: 79). In other words, they would retain their value even if we were unable to detach ourselves from aspects of them, that is, even if we were dependent on them as far as those aspects were concerned.

both.⁵³ It would be, I believe, a grave error of moral and political judgment to maintain that this is a blessing and not a curse. So, I think Kimel shows acuity of judgment when he says that “*dependence* on personal detachment is itself not so much a thing of value but a predicament” (2003: 142). But this, alas, contradicts his claim that personal detachment remains a good thing when examined in isolation, that is, even in a world where we are dependent upon it—as would have to be the case if the ‘good’ claimed is non-instrumental. As we have seen, the value of personal detachment differs from the value of personal relationships in this respect, which would, it seems, reflect that the latter but not the former has non-instrumental status.

This is why I believe the first thought in the above passage is confused, and can be safely disregarded. The emphasis on the value of the *option* of personal detachment, as opposed to relations of personal detachment *simpliciter* does, however, bring us closer to the second thought. Now, of course, if all this means is that personal detachment is non-instrumentally valuable, but somehow only when we are not dependent on it, then, as we have just seen, it will be a nonstarter too. But this is not the only way to unpack the value of the option. There is also the possibility that when personal detachment and personal relationships coexist side by side as options, their values mutually reinforce each other. Once we recognise this possibility, another route to the non-instrumentality of the value of personal detachment presents itself, namely, that having it as an option is constitutive of the enhancement of the value of our personal relationships, making them deeper or more meaningful in some *sui generis* way. This is the second thought in the Kimel passage, and it is, to my mind, far more penetrating than the first. I believe it puts a finger on something primordial in the doctrine of liberalisation.

To be clear: the proposal is that the option of personal detachment constitutively enhances, deepens, or otherwise makes more meaningful our relationships, and in that way receives a non-instrumental significance. What exactly does Kimel have in mind? His example (2003: 79) of the liberalisation of family relations is indicative:

⁵³ This is the point of an anecdote from Patricia Williams (1991) explored in the closing pages of Kimel (2003: 139-42). Williams recounts how she, a New York lawyer, and a black woman, cannot have confidence in a tenancy agreement with a white landlord without appeal to the mechanism of contractual enforceability, whilst her white male colleagues both can and prefer to proceed in an informal manner on the basis of trust, or, as she says, with a “handshake and good vibes” (1991: 146).

even where examples such as family relations are concerned, the relative freedom to draw lines, to exert control over the scope, the depth and the intensity of the relationship usually tends to improve the quality of the relationship and enhance its value for the parties, rendering it more meaningful—indeed, it is tempting to say more viable—than the kind of all-encompassing family ties that deny the parties the opportunity to maintain relative privacy or detachment in any way or with regard to any aspect of their lives.

The general idea here seems to be that our relationships become more meaningful to the extent that we gain control over the scope, depth and intensity of what is required of us within the sphere of the relationship. Control comes, fundamentally, from having the option to keep certain of our thoughts, feelings and actions private, thus cordoning off those aspects of our lives from the demands of the relationship. These aspects of our lives are often important, and so we do not have this control if we are simply unable to realise the relevant thoughts, feelings and actions altogether. Instead, we have to be able to do this in a manner that suits us *qua* individuals, and not just as required under the jurisdiction of the relationship, where the other can always interfere. With thought and feeling, whether we have this option basically depends on the dispositions of the parties: how pushy or violent they are; how poker-faced and resilient we are. But with action, it basically depends on how things external to the parties are set up. This is in part a function of there being favourable social and cultural norms that permit us to detach ourselves from our relationships insofar as they require us to do things in a certain way. But it is, in equal measure, a function of there being concrete ways to do these things in a manner that suits our individuality, which there would not be but for the existence of markets for services in these things.⁵⁴

⁵⁴ This explains why the fact that online spaces like Reddit offer personal detachment in the absence of contractual relations is not an objection to this reading of Kimel's theory. The non-instrumental benefits Kimel proposes cannot be had without markets for services, for which contractual relations plausibly *are* an essential ingredient, in a way that online spaces simply are not. The personal detachment such spaces afford is plausibly of merely instrumental value, if it is of value at all.

It is this last idea that will concern us more specifically. Our relationships are said to be more meaningful when we have the option to detach ourselves from them in the manner just described, giving us control over how we do the important things we have to do within their sphere. As there comes to be a market for more and more human activity traditionally thought to fall within the realm of our personal relationships, our options for outsourcing them to service-providers become more and more extensive, running the whole gamut of human activities, from the most to the least tangible. Our having these options, whether we use them or not, is supposed to deepen the meaning of our relationships in that it becomes a constituent of their non-instrumental value that their shape is sensitive to our individuality. That we could, if we pleased, opt to proceed in a more detached way in any aspect of a given relationship seems to be part and parcel of the depth of its value to who we are as individuals.

I believe this proposal has a lot going for it. The idea that our personal relationships are deepened by having options over the extent of their jurisdiction in our lives chimes keenly with our modern sensibilities. It is also plausible that the enhanced sensitivity of our relationships to our individuality are constituted by the existence of the option of personal detachment, and that this option, in turn, cannot be had without markets in services and their contractual foundations. Does this interpretation of Kimel's appeal to personal detachment secure the inflationist ambitions of his theory? I want to argue that, despite appearances, it does not. Even on this promising reading, I submit, Kimel faces political challenges that thwart the inflationist ambitions of his theory.

6.3.2.B. *The Politics of Personal Detachment*

Think of Kimel's proposal in the following terms. There are, first of all, what we could call *old-style* relationships bound up with norms and expectations—quite possibly both asymmetrical and systematically so across morally arbitrary groupings of parties—that are expressive of traditional values and authority. The potential meaning for us of these old-style relationships is genuine.⁵⁵ But it is, in a second step, deepened by our having

⁵⁵ We may, as feminists, instinctively take this claim to make Kimel's proposal unworthy of sustained attention. This seems to me too quick, as there are serious feminist intellectuals who have, in good faith, made this claim explicitly. Even before her rightward turn, Jean Bethke Elshtain writes, for example: "Rather than denying women the meaning their traditional world

the option of personal detachment, which turns them into *new-style* relationships. Since the relevant deepening on display—viz. their sensitivity to our individuality—cannot be achieved without our having the option of detachment, the option is constitutive of the deepened meaning of the new-style relationships.

Before anything, we should register the clear emancipatory impulse behind Kimel's proposal. Whatever else we think of it, the proposal begins as a reflex against traditional authority, and should be commended as such. With this firmly in place, I will begin to apply some pressure. Notice that, on Kimel's proposal, new-style relationships are going to turn out non-instrumentally worthwhile just so long as the degree of detachment they exhibit (which could be nil) reflects a sensitivity to the individuality of the relevant party. I want to problematise this implication. In particular, I will consider a counterexample to it that, fully spelled out, reflects a certain tension in the politics of the wider proposal.

Here is the counterexample. Imagine two people for whom career is the organising principle of life, who are intensely aspirational, ambitious and status-driven, and who fundamentally identify with the values these traits express. They have heard that there are certain advantages that come with romantic involvements, and they find in each other precisely the degree of involvement their respective career goals will tolerate. At some point, they hear that kids, too, have their advantages, and so they decide to become parents. When they do, of course, it turns out that being parents is rather more work than their careers will allow. And so, slowly but surely, they outsource *every last* aspect of the material, social and emotional labour of care for their children to a rotating circle of third-party providers. In the parents' eyes, the relationship with their kids is ideal: its non-existent demands are exactly sensitive to who they are as high-powered careerists. The thought is that, having outsourced so extensively, they surely cannot honestly count a meaningful relationship to their children amongst the sources of non-instrumental value in their lives.⁵⁶

provided, even under conditions of male domination, feminists should move to challenge a society that downgrades female-created and -sustained values" (1995: 268).

⁵⁶ It is important to stress that, in the case at hand, the outsourcing extends to literally every last aspect of the material, social and emotional labour of childcare. Intuitions may, of course, and rightly, differ for parental practices of outsourcing childcare that are less extensive than this in any way. It is also important that the parents are 'outsourcing' in the specific sense that they contract out their childcare to a third-party provider for a fee. Perhaps the intuition goes away if we imagine similarly extensive 'outsourcing' of childcare to third party providers who do not

There is a certain kind of conservative that could, I think, agree with a certain kind of progressive about this counterexample. For example, some social conservatives will cite the sanctity of the family and its social bonds, which they regard as the foundation of society and social order. More often than not, this comes bundled with the claim that it is unnatural for parents not to care for children themselves, perhaps with approving gestures towards a gendered division of labour alleged to reflect biological differences. Of course, such grounds for suspicion of the non-instrumental value of the relationship need not worry Kimel, for, as we saw, his proposal is intended as an emancipatory reflex against just such a conservatism. A fundamental disagreement with traditionalist values is baked into the proposal from the start.

More worryingly for Kimel, there are serious progressives who will regard this as a counterexample, and indeed on emancipatory grounds. The idea would be that at least some of what is outsourced here is the stuff of the good life, with the result that we are *alienated* from the relational sources of our own good.⁵⁷ Our relationships, which should be a source of deep meaning and significance, instead come to confront us as something alien and external. The activity of satisfying their demands is reduced to the status of services, which, by contracting out, facilitates our pursuit of something yet further that is the true source of meaning in our lives: work, wages, and the freedoms that come with socio-economic security. When our social relations become alienated in this way, the

expect a fee. This probably depends on what exactly is being imagined. Is the arrangement like an extended kinship or community group? Or is it more like how Engels thought things would be organised after the revolution: “Private housekeeping is transformed into a social industry” (1975: 139)? Or perhaps the very distinction between parents and third parties is imagined away, as in Alison Jaggar’s socialism, which “would become the basis for entirely new forms of social and sexual groupings” (1983: 226). The insistence on contractual childcare provision is meant to sidestep these complex issues.

⁵⁷ In using the term ‘alienation’ to label the phenomenon described, I follow the early Marx, who recognised several ways in which wage workers could be ‘alienated’ or ‘estranged’: (i) from their product, (ii) from their work activity, (iii) from their human potentiality, (iv) from other humans (1988: 71-9). The charge considered here is that outsourcing in this manner described alienates us from our social relations in sense (iii). As Marx would put it, the “labor, *life-activity*, *productive life*” in which our social good is realised “appears to man merely as a *means* of satisfying a need—the need to maintain the physical existence. [...] Life itself appears only as a *means to life*” (1988: 76).

thought goes, we suffer a form of domination or oppression.⁵⁸ And this, I think, should worry Kimel. Though his proposal sets out to resist oppressive structures of traditional authority, it inadvertently ends up valorising oppression in another guise: the alienation of our social lives under the pervasive imperatives of the market.

Nancy Fraser would say that the counterexample shows that Kimel's theory suffers the same political fate as did many emancipatory projects of the latter 20th century. The starting point for Fraser is *The Great Transformation*, Karl Polanyi's seminal 1944 study of market society. In Polanyi's analysis, market forces tend to dissolve social bonds, so that, as the economy gets increasingly marketised, ordinary people instinctively resist the attendant threat to social life as they know it. As a result, societies whose economies assume a market form are constitutively politically unstable. Polanyi calls their powerful political dynamics the "double movement" (2001: 80-1), using it to explain apparently heterogeneous historical outcomes ranging from the establishment of social protections (e.g. welfare systems) to the rise of fascism. For Fraser, however, the analysis conceals as much as it illuminates, and will continue to do so until we recognise—opposed *both* to the forces of social protection *and* marketisation—the distinct forces of emancipation. We must extend the analysis from a 'double' to a "*triple movement*" (2013: 230).

To see why, take Fraser's analysis of the fate of feminist critique of social insurance schemes based on the male-breadwinner model of the 'family wage.' The original intent behind the critique is clearly emancipatory: the family wage model is patriarchal in that it entrenches the second-class status of women in the economic and social spheres. But without a clear view of the triple movement, Fraser suggests, subsequent feminists often struggled to frame their emancipatory opposition to social protectionism in ways that did not unwittingly play into the forces of marketisation. As she puts it (2013: 220-1),

Our critique of the family wage now supplies a good part of the romance that invests flexible capitalism with a higher meaning and a moral point. Endowing their daily struggles with an ethical meaning, the feminist romance attracts women at both ends of the social spectrum: at one end, the female cadres of the professional middle classes, determined to crack the glass ceiling; at the other end,

⁵⁸ As Philip Kain puts it, "One can be dominated and oppressed without being alienated. But if one is alienated, one is certainly dominated and oppressed" (1993: 122).

the female temps, part-timers, low-wage service workers, domestics, sex workers, migrants, EPZ workers, and micro-credit borrowers, seeking not only income and material security, but also dignity, self-betterment, and liberation from traditional authority. At both ends, the dream of women's emancipation is harnessed to the engine of capitalist accumulation. Thus, second-wave feminism's critique of the family wage has enjoyed a perverse afterlife.

Just as feminism falls prey, in this way, to what Fraser elsewhere calls the "cunning of history" (2013: 211), the worry for Kimel is that what starts life as an emancipatory proposal against traditional authority ends up being co-opted by the oppressive forces of marketisation. That is the force of the counterexample. Is there any way for Kimel to respond?

An immediate thought might be that sensible liberals must have things to say about the value of autonomy that can put *legal* limits on the freedom to contract sufficient to rule out the counterexample. I think that, on reflection, sensible liberals are unlikely to find this solution attractive, and Kimel (2003: 129-34) is one of them. On his view (2003: 131), "An autonomous life is valuable when spent in the [...] pursuit of valuable activities and relationships, but not otherwise."⁵⁹ Personal autonomy is thus not just a matter of having a range of any old options, but specifically valuable ones. To protect autonomy, a liberal legal order may thus eliminate options that are reasonably regarded as bad,⁶⁰ including options that are bad because they are such as to essentially threaten autonomy: they are exploitative, unfair, or otherwise unduly restrictive. But the contractual options in the counterexample are not eliminable by this route. Contracts for childcare services are unobjectionable, at least taken one by one. Of course, it may be that they threaten autonomy merely accidentally, for example, if their wanton use makes it harder to participate in meaningful parent-child relationships. In that case, it may be possible to limit these contractual freedoms. But this threat to autonomy only obtains if the availability of such contracts tends to hamper meaningful parent-child relationships *in general*, which need not be the case: the careerist parents, we may assume, are quite

⁵⁹ This view of personal autonomy has been defended by Joseph Raz (1986: 378-81).

⁶⁰ In Kimel's words (2003: 132n50), the argument, roughly, would be this: "the autonomous choice of bad options *is* in fact morally *worse* than the non-autonomous choice of bad options, and worse precisely because it is an autonomous choice."

idiosyncratic in their outsourcing practices. So, for sensible liberals, as Kimel says (2010: 230-1), the focus is always on “what kind of transactions the law of contract should facilitate,” so as not to actively “facilitate immoral conduct.” For unobjectionable kinds of transactions, a sensible liberal cannot limit our freedom to make them simply because we could opt to squeeze them into our lives in silly ways.

Another response for Kimel might be as follows. In general, it may be that there is value in our having certain powers even as particular exercises of those powers are not themselves valuable in that way. Thus, having the option of detachment may really be worthwhile in that it potentially makes for more meaningful relationships, even as the use of that option fails to do so in a given case. In our preferred language, this suggests a type-token ambiguity in the new-style relationships Kimel proposes. In saying that the option of detachment makes for more meaningful relationships, what he means, he can claim, is that it makes for a more meaningful *type* of relationship: the new-style type of relationship. But as with any worthwhile type of relationship—friends, family, lovers, you name it—its *tokens* may not themselves be worthwhile, because they are in some way pathological tokens of the type, i.e. the result of bad or pathological exercises of an otherwise worthwhile option of detachment. The suggestion would then be that the counterexample is cherry-picked from the defective tokens of a generally worthwhile type of relationship, so that, although the token parent-child relationship in question really is sensitive to the individuality of the parents, Kimel is nonetheless not forced to concede that it is in fact worthwhile.

But if this is the idea, then Kimel needs to tell us what exactly he thinks it is about the parent-child relationship in question that plausibly suffices to make it a pathological token of the type, and so not in fact worthwhile. It cannot be that the degree of detachment it involves fails to reflect the individuality of the parties, since it does not fail to do this *ex hypothesi*. Another suggestion could be that the extent of detachment in the relationship amounts to a form of abuse or neglect—a suggestion that becomes more plausible when we focus our attention on the kids. But is it really so clear that the kids suffer abuse or neglect in the relationship? The kids are, after all, being cared for by someone, and that care is ultimately paid for by the parents. Indeed, the kids may, having been brought up the way they have, feel that their relationship to their parents is just as they wish. At this point, perhaps, we may want to attribute something like false consciousness to some or all of the characters involved. Yet by Kimel’s own lights, this

solution is unavailable: it would presumably be anathema to the liberal emphasis on individuality within his wider proposal.⁶¹

This train of thought suggests a more promising line for a liberal like Kimel to take. He could reply that the intuition behind the counterexample—i.e. that the relationship is alienating and oppressive despite the fact that it reflects the parties' individuality—is based in a comprehensive doctrine about the good life. He can then say that, though he respects this doctrine, which seems reasonable enough, he objects to the tacit assertion that it is ultimately correct. This may be on grounds of liberal neutrality: that trying to elevate this doctrine above those of the characters in the counterexample is patronising, paternalistic, or otherwise fails to treat them with equal respect and concern. But it need not be. The objection could just as well be on the epistemic grounds that the superiority of one comprehensive doctrine over any other is unknowable, which may be, in turn, for the metaphysical reason that there is no supreme value against which conceptions of the good can be meaningfully compared. An objection on any of these grounds could, it seems, be enough to get Kimel off the hook.

No matter the ground of the objection, however, the promise of this approach turns out to be illusory. The point of the counterexample is to present a new-style relationship that is intuitively not, as Kimel would have it, non-instrumentally worthwhile. The reply at issue objects to the counterexample's tacit invocation of a substantive conception of the good. But if Kimel really can nullify the counterexample in this way, he must also, by the same token, nullify his own proposal that new-style relationships are just the old-style ones but with a *deepened* meaning or non-instrumental value. This is because it is far from clear how Kimel can justify his claim to deeper meaning—where that is a way of being *more* worthwhile—without invoking some comprehensive doctrine himself, thereby falling foul of the same objection, whatever its ground.

Perhaps at this point Kimel will concede that he misspoke in using the language of 'deepening.' What he meant, he might say, is that the axiological upshot of our gaining the option of detachment is not so much a deepening of the old-style relationships, as their creeping replacement by mutually exclusive new-style ones. It is not that the new-

⁶¹ As Meir Dan-Cohen (2002: 18) would say, we attempted to show the token relationship to be pathological, first by attributing "*alienation*" to the parents, then by attributing "*bad faith*" to the children. The difficulty is that, to succeed, we need a suitably neutral criterion for the relationship to count as a pathological token of the type, which does not seem to be forthcoming.

style relationships are arrived at by any sort of improvement of the old-style ones, so that the latter come to have a greater non-instrumental significance for us. It is rather that the values embodied in the old- and new-style relationships are just fundamentally different, such that it is simply not possible to meaningfully compare the ways our participation in them can enrich our lives. There is no sensible answer to the question: which is the better, richer, or deeper style of relationship? It is just, as they say, apples and oranges.

Of course, if this is right, then the often fraught political contestation around these issues in recent times will turn out to be basically pointless. Since the traditional values embodied in old-style relationships are just some amongst many incomparably diverse values, each worthwhile in their own way, it turns out to be little more than a matter of taste whether, for instance, the place of women in economic and social life should be to furnish the background conditions for the productive activity of men. Similarly, since the values of liberal individuality embodied in each and every new-style relationship are just some amongst many incomparably diverse and equally legitimate values, it turns out, once again, to be no more than a matter of taste whether we should aspire to a form of life that does not cultivate social alienation. Perhaps the voices of social conservatism, laissez-faire and emancipation are all, in the end, just talking past each other. There is a type of liberal who could find comfort in this thought, though I am not convinced Kimel would regard himself as one of them. But this, it would appear, is the price he has to pay to salvage his appeal to the distinctive value of personal detachment.

6.3.2.C. *Contract as a Form of Technology*

To briefly recap. According to Kimel's pessimistic inflationism, the non-instrumental significance of contract lies in the value of personal detachment. Specifically, the idea is that, without it, we would not have the option to proceed in our special relationships in a way that reflects our individuality, or the deepened meaning said to come with it. This view, though attractive, faces a counterexample that can only be avoided by retreat to a form of value pluralism that empties it of its emancipatory potential.

Of course, this is bad news for Kimel's inflationism. But it may strike us as bad news quite apart from that debate. After all, it is plausible that contracts constitute our having the option of personal detachment, and that this really can be worthwhile in the manner

proposed, i.e. that it can make our relationships more meaningful by giving us control over the extent of their jurisdiction in our lives. We may reasonably not want to give up on these ideas. Can we salvage Kimel's appeal to personal detachment in contract theory in its original emancipatory spirit?

The discussion of 6.3.2.B brought out several value-theoretic desiderata that any such attempt would have to meet. To start with, for old-style relationships to be such as to have their meaning deepened by the option of personal detachment, (i) they must genuinely have had a meaning to deepen, at least once upon a time. But here and now, in the twenty-first century, where we are lucky enough to have the option of detachment, old-style relationships can no longer be genuinely worthwhile: they smack of traditional authority, are outdated, and are—in a word—*old*. As such, (ii) it must be that only new-style relationships can be genuinely worthwhile. Nonetheless, (iii) it must also be that not just any old exercise of the option of detachment necessarily makes for a genuinely valuable new-style relationship. Otherwise we will have to admit that where detachment proceeds in an unthinking or excessive manner, as in the counterexample, the alienated relationships that result are worthwhile in their own way too. Desideratum (i) falls out as a presupposition of how personal detachment is said to receive its significance for us, whereas (ii) and (iii) are needed to avoid the value pluralism discussed before, and the political impasses therein contained.

I would like, in closing, to explain how we might get results (i)-(iii) out of an appeal to personal detachment in contract theory. Admittedly, the story will be impressionistic. In my defence, it is not really my job to supply it. I offer it in a conciliatory spirit, a way to preserve what is attractive in a view I have otherwise been concerned to argue against. We will see that it makes no difference to the deflationary ambition of the paper whether my story is persuasive or not. The present context is one in which something—even something fairly sketchy—is better than nothing.

My story takes shape via reflection on the relations between values and *technology*. We have been understanding noninstrumental values as the sorts of things that we are able to benefit from directly by engagement with them for their own sake. Suppose, with Joseph Raz, that specific noninstrumental values are (a) partly constituted by attendant forms of excellence and the standards of successful engagement implicit in these (2003: 31-2), which (b) are as a rule ultimately dependent for their existence on social practices that do or did once sustain them (2003: 19). These assumptions about value allow us to

make the following structured observations about the potential impacts of technology on values and our engagement with them.

The introduction of a new technology into society will have many effects, some of which touch on values and our engagement with them in various ways. Most obviously, there will be intended effects, at least if it is a good technology that actually achieves the purpose it was designed to fulfil. What this looks like for specific technologies will vary from case to case, but it seems plausible that, in general, the purpose of technology is to enable us to better do things that facilitate our engagement with what is of value. When a good new technology becomes available, it makes excellence in our engagement with valuable things more attainable, at least initially. On top of these intended effects, there will be unintended effects too. Some of these relate to the technology's achievement of its purpose, and others not. We will focus on just those effects of new technology insofar as it is good at what it was designed to do, which, as we have just seen, will include both intended and unintended effects.

My claim is that the intended effects of introducing good new technology will have unintended effects on the very value that it succeeds in helping us engage with. Consider the impact of artificial intelligence (AI) on chess. Chess has been a great game ever since it was invented. But it appears that the introduction of neural net AI engines is changing the value of chess at the elite 'super' grandmaster (GM) level. Just like any technology, the purpose of AI chess engines is to enable us to better do things that help us engage with what matters—in this case chess. They do this very well: they offer deep evaluations, calculate the best lines, show us what (not) to do in a given position, giving a rationale, and so on. All of this helps us to play excellent chess. AI engines have become ubiquitous in the chess world for precisely this reason: every chess player these days has access to one. At the super GM level, high-end engines are available to help gain an advantage in classical chess, a format that remains prestigious within the chess world. As developers compete to write better engines, super GMs stay up-to-date to maintain their edge. As the limits of AI engines—or at least human players' ability to assimilate their insights—become clearer over time, super GMs and their teams have begun to experiment with how to integrate the technology into training for the best results in classical chess. These are some of the unintended effects of the success of chess AI engines at fulfilling their purpose. They are effects on the social practices that sustain the value of chess.

As the social practices that sustain the forms of excellence in an activity change, so too, I claim, does the value that those forms partly constitute. In particular, what counts as excellence in how we engage with the relevant value comes to include standards of propriety in the application of the new technology. To see this, consider the practices of super GMs and their teams today. Having the players use AI as much as possible—for example by trying to memorise as many top AI lines for as many moves of as many openings as possible—has not emerged as a promising way to gain an advantage over-the-board. What works better is for the players' teams to research themes and concepts based on AI analysis, to be humanly explored—including for theoretically weak lines that harbour opportunities for a human opponent to make mistakes—by players who themselves studiously avoid contact with AI engines.⁶² In other words, what it means to play classical chess at the super GM level has come to include the sophisticated use of AI; and *that*, I submit, makes the value of chess today fundamentally different from the chess of old. Now that we have AI, there is a clear sense in which a super GM who pines after the old chess would be missing the point of chess as it exists today.⁶³ This is not to deny that chess back in the Golden Era was a great game. It is to say that super GMs (or their teams) who systematically shun AI simply fail to grasp what is good about the only

⁶² Peter Heine Nielsen, who coaches (at the time of writing) world number one Norwegian GM Magnus Carlsen, describes their process thus: “Magnus might feel a move gives him the best chances, puts his opponent under awkward pressure, even if the computer might discard it. [...] Modern chess is like this: If we all follow what the computer tells us in analysis, we both have the same source, the same ideas, and the game will be a draw. So we try looking for on-the-edge concepts... That's chess. That's sport.” (Available at <https://www.techopedia.com/magnus-carlsen-how-intuition-and-ai-shape-the-best-chess-player-in-the-world>, accessed June 17th 2024.)

⁶³ Perhaps one of these would have been the great American GM Bobby Fischer. Even before powerful chess engines were widely available, Fischer lamented the proliferation of chess theory, which he believed was responsible for the death of the chess of his heroes, turning a game of spontaneity and creativity into one of memorisation and preparation. He famously advocated the shift to a randomised variant of chess called ‘Fischerandom’ or ‘Chess960’ to fix this, saying, “I invented Fischerandom chess to keep chess going. Because I consider the old chess is dying, or really it's dead. [...] I want to keep the old chess game. But just making a change so the starting positions are mixed, so it's not degenerated down to memorisation and prearrangement like it is today.” (Available at <https://chess960.net/chess-quotes/>, accessed Mar 20th 2025.)

chess that exists any more. The value of the chess of old has not just been replaced but *superseded* by that of the new chess.⁶⁴

We can understand how the new availability of the option of personal detachment interacts with the value of our personal relationships in an analogous fashion. Perhaps once upon a time it was possible for us to find meaning in old-style relationships, just as chess really was great in the Golden Era. But when we acquired the option of personal detachment and the associated control over the reach of our old-style relationships, this really did enable us to better have lives fit for ourselves as individuals, as was the express purpose of our having the option. As more and more people came to exercise the option, the social practices sustaining the value of old-style relationships changed, transforming that value. What counts as doing well in this new style of relationship—the attendant forms of excellence—has come to include standards of propriety in how we exercise the option of detachment.

Notice how this analogy satisfies desiderata (i)-(iii). It straightforwardly satisfies (i), since it allows that old-style relationships were genuinely meaningful once upon a time. Moreover, just as in chess, the value of new-style relationships is said not only to replace but to supersede that of the old. So the analogy also satisfies (ii). Now that we have the option of personal detachment, *only* the new-style relationships can be truly worthwhile. There is really something tragic about those social conservatives who are nostalgic for personal relationships expressive of traditional values: in refusing to get with the times, they are condemned to miss the point of the only style of meaningful relationship that exists for them to participate in here and now. And finally, (iii) is satisfied, because the analogy suggests that figuring out how to integrate the option of personal detachment into our relationships to secure excellence in the new style is likely to involve sensitivity, thoughtfulness and experimentation. Just as the wanton use of AI is not a way to excel at the new chess, we should not expect that wanton detachment will be a way to excel in our personal relationships as they exist today. Those high-powered professionals who

⁶⁴ If the value of the new chess supersedes that of the old chess, the latter becomes redundant. That is not to say that there can be no historical value in imaginatively projecting ourselves into the social milieu and practices of the past. The vivid reconstruction of old values can help bring history to life, deepening or complicating our self-understanding. But such historical value is not the same as the old values themselves.

choose to hollow out every aspect of their relationship to their kids need not be doing a good job as parents in the new style.

Accepting a story along these lines allows us to salvage the emancipatory appeal to the value of personal detachment in contract theory. We can say that, without contract, we would not have the option of personal detachment, and we would thus be deprived of the deepened noninstrumental significance of personal relationships as they figure in our lives today. And we get to say this without conceding to social conservatives or proponents of *laissez-faire* that whether we prefer our style of relationship over theirs is, in the end, no more than a matter of taste. But to get the story going—and this is the point—we had to analogise the option of personal detachment with AI engines in chess. Of course, this presupposes that, like AI, the option of personal detachment is a form of technology and is, as such, first and foremost of *instrumental* value.

The consequences for the contract-theoretic appeal to personal detachment should be clear. Recall how that appeal is supposed to work. Contracts are said to be an essential constituent of the value of the option of detachment, deriving their own value from this source. If that is so, then, on the present story, the value of contracts will of course turn out to be ‘merely’ instrumental. To reinvigorate the emancipatory potential of Kimel’s contract theory, we ended up telling a story that *deflates* it.

6.4. Concluding Remarks

The broadly liberal tendency towards inflationism in contract theory says that, without contracts, we would be deprived of certain noninstrumental values, so the contribution contracts make to human life is itself noninstrumental. Section 6.3 has been a sustained attempt to show that no inflationism—whether based in the Enlightenment-optimist values of *doux commerce* (6.3.1), or the more pessimistic value of personal detachment (6.3.2)—looks to be especially promising.

In the context of the chapter as a whole, the point of this is to remove the primary obstacle to accepting the moral-philosophical analysis of contractualisation I offered in section 6.2. The analysis proceeds on the basis of a deflationist theory of contract as an instrumental relationship (6.2.1). It aims to show how some of the distinctively human ills that attend the contractualisation of our social relations exemplify the politics of

tragedy (6.2.2). To the extent that my analysis is persuasive, the chapter plays its proper role as outlined in the preamble to Part Three.

I do not imagine for a moment that there is not plenty of room left for disagreement over the details of my deflationist contract theory. And I recognise that, even accepting my deflationism, we may still question the ethical consequences I claim this has for our social relations under the pressures of contractualisation. There may be good reasons to be left unpersuaded that contractualisation really does in the end exemplify the politics of tragedy, as I have suggested. But I hope to have shown that, if there are good reasons, then a prior commitment to inflationism in contract theory is not likely to be one of them.

Security, Social Relations, and The ‘Political’

7.1. Our Security Interest

In the last chapter, I analysed the contractualisation of our social relations, showing how its distinctively human ills exemplify the political dimensions of tragedy. In this chapter, I focus on how the norms of our social relations necessarily distribute security zero-sum between the parties, making them susceptible to distinctively political considerations. My emphasis here will be less on how we collectively understand our social relations, and more on how we practically shape them.

Let me start by saying what I mean by ‘security’ in this context. In Chapter Three, I suggested that what it is to lead a life is to engage with the things that matter to us whilst respecting those that do not. This entails that we have no life we are leading unless we exert ourselves to meet the demands of what matters: were we never to make an effort, we would never do so better or worse, and so our lives would be contentless. Our efforts are, as it were, the fundamental medium through which our lives are led. Note however that the efforts we exert are not identical with our biological existence. We surely need to be alive to have a life we are leading, but we also need, in addition, to exert ourselves through action. Equally, to have a life we are leading cannot consist solely in exertion: sleep, for example, is required to be alive, though it is effortless is we can achieve it. Even

our waking life is plausibly not led well in constant exertion: if we call the absence of exertion in waking life 'leisure,' then a life well-lived plausibly cannot be had without sufficient leisure or relaxation. But at least some of our waking life must be spent making an effort, if our life is such as to go better or worse.

Though leading a life takes more than just being alive, it does require our being alive, and that, I believe, accounts for the finitude of our efforts. The temporal, spatial and psychological boundaries of our biological existence ensure that we only have so much effort we can reasonably be expected to give. I will assume that how much effort we can reasonably be expected to give is one and the same for us all. Of course, some have the grit to spend more of their waking lives making an effort than is reasonable, or indeed, to this end, to force more of their lives to be waking. Some have the savvy to make their efforts go further than can be reasonably expected, by all manner of psychological, social and technological means. But efforts made through grit and savvy, as I understand them, go beyond the efforts that can be reasonably expected of us, which we all possess in the same finite degree. By 'efforts,' in what follows, I mean those that are reasonable, unless otherwise qualified.

Due to the finitude of our efforts, the demands of what matters may be contingently mutually unsatisfiable by us, so that we are torn between them. In this way, our worldly condition can leave our efforts spread so thin that we cannot but fail to live up to some of the demands we face, threatening the quality and integrity of our lives. Our efforts are spread too thin when there is no way we can reasonably exert ourselves without falling short of the things that matter, making our lives go worse thereby. This possibility, built into our condition, perhaps explains the perennial impulse to withdraw from the world. We may seek to extricate ourselves from our attachments, or to isolate ourselves entirely from the things that matter, in a bid to reduce the risk of being spread too thin by their competing demands. Turning inwards may serve, to a degree, to alleviate the strain that the things that matter place on our dwindling reserves of effort. But it is a sorry strategy because, at the limit, it alleviates us of our lives as actually led, reducing us to a vegetative existence. To have a life we are leading, we must run the risk of being spread too thin.

I believe this establishes that, in general, we have an interest in protecting our finite pool of efforts against being spread too thin. What decreases the likelihood that our pool of efforts will be spread too thin advances this interest, and what increases the likelihood sets it back. The interest I have in mind comes apart from our interest in wellbeing. On

the one hand, what protects our finite pool of efforts need not make our life go better thereby: to put it crudely, we may have plenty efforts left in the tank without using them very well, or indeed at all. On the other hand, our wellbeing consists in responding non-instrumentally to what matters, which necessarily takes an effort, thereby dwindling our finite reserves. Our interest in protecting our pool of efforts is therefore focused not so much on living up to the demands of life, as on not being overwhelmed by them. When our efforts are spread too thin, and we are overburdened, we cannot but fail to live up to the demands of life, rendering the quality of our lives precarious to that extent. In this sense, and for want of a better expression, I will label the general interest we have in protecting our finite pool of efforts *our security interest*.

Our security interest has material and psychological aspects. To protect our finite pool of efforts is to protect our ability to live up to the demands of the lives we are leading, which is a resolutely material affair. But if our security has this material aspect, it has a psychological aspect too, owing to the downstream repercussions of falling short of life's demands. Falling short not only makes our lives go worse to that extent; it calls for efforts to repair the damage we have done, and feelings of agent-regret for what we have so far left unrepaired.¹ So, when we are spread too thin, and cannot but fall short in one way or other, we open ourselves up to painful feelings that we target upon ourselves. As such, protecting our finite pool of efforts against the vicissitudes of life protects us against the darker shades of moral emotion, against an inner life whose tendencies are more likely to be destructive. I believe this makes our security interest as much a psychological as a material interest.

Quite generally, then, our worldly condition—and the security interest it implies—lends a certain ambivalence to our relationship with the things that matter, making them at once a source of value and insecurity in our lives. In this chapter, I want to zoom in on this ambivalence as it arises for our valuable social relations. Seen through the lens of our security interest, it is clear that there are distinctively political dimensions to the value embodied in the constitutive norms of our social relations—or so I claim. Section 7.2 explains how some of these norms necessarily distribute security (and so insecurity) between the parties in a zero-sum fashion, so that considerations of distributive justice necessarily apply to how the parties practice those norms. Section 7.3 explains how the

¹ See the preamble to Part Two and Chapter Three for defence of this claim.

fact that people stand in several social relations at once allows for a distinctively ‘political’ subset of such considerations. Section 7.4 explains how these considerations can lead us to practice the norms of our social relations with a view to how tragedy gets distributed across the wider social order. Section 7.5 briefly concludes.

7.2. Local Distributions of Security

Before we begin, some preliminaries about the concepts of social order, social relations, norms and practices. The picture that follows is schematic but hopefully not implausible, and is intended, at least in broad strokes, to reproduce and extend upon the one we find in Joseph Rouse (2023).

It is plausible that social orders are at least partly, but essentially, constituted by the social relations that obtain between people in the societies that are ordered thereby. To see this, take the situation from the Irish context discussed in Chapter Two. An example of a social relation, as I understand it, would be that which obtains between the landlord of the Dolphin House estate, Dublin City Council, and the tenants of that estate. In that particular situation, it so happens that the landlord is also a local authority, which stands in a separate social relation, *qua* local authority, to the Irish central government. Equally, amongst the Dolphin House tenants, there are likely some who have dependent children, to whom they stand in a separate social relation *qua* parents. The prevailing social order in the Irish context, I suggest, consists in how the individuals subject to it stand in some constellation of social relations like the ones mentioned. These social relations, I claim, essentially constitute the social order insofar as Irish society, and its members’ patterns of reason-governed behaviour, would be organised differently if they were bound to each other in different social relations.

I count the relations between central and local authorities, landlord and tenant, and parent and child, amongst ‘social relations’ in the relevant sense. In this usage I follow, for example, John Gardner (2018b: 36-7) and Joseph Raz (1977: 228) before him. To preempt any confusion, by ‘social norms,’ ‘relations’ and the ‘orders’ they constitute, I do not refer to normative phenomena that presuppose the institution of law, although they may have their legal counterparts. It may seem implausible that there could be extra-legal analogues standing behind familiar legally constituted relations, as between central

and local authorities or between landlords and tenants. On the contrary, I believe that some such distinction is needed to explain the way the law develops over time. Consider the liberalisation of the norms governing marital relations in English law. No doubt social agitation over the last few centuries—people’s changing social mores and norms—played their part in shifting the law of marriage to where it is today. Equally, incremental shifts in law over that period, both in statute and precedent, will have served to reinforce and set the direction of travel for people’s social practices in their marital relations. I am not sure how to explain either causal mechanism unless legal norms and relations have their extra-legal, social analogues.

In what way do social relations ‘bind’? For any given kind of social relations, I claim, there will be certain directed obligations and correlative rights that parties to it must be subject to, if they are to count as standing in that kind of social relation. Call these the *constitutive norms* of social relations of given kinds.² A ‘government’ with no obligations whatsoever to its people is no government at all, nor a ‘landlord’ with no obligation to maintain their tenants’ housing, nor a ‘parent’ with no obligation to help raise their children. For constitutive norms to exist—and with them the kinds of social relations they partly constitute—a sufficient number of people have to pattern their behaviour so as to uphold the norms as if there were something at stake in the kind of social relations those norms constitute. They do this by conforming with the norms and giving them their due in deliberation, as well as by holding themselves and each other accountable if they fail to do so. I will put this by saying that norms do not exist unless they are actually practised. We can describe social relations of any arbitrary kind by specifying the norms that parties would have to practice, to be counted as standing in those social relations. But if no, or not enough, people actually practice the norms, then neither the norms nor the social relations they describe can be said to have any practical existence.³

² See Chapter Four for more on the explanation of constitutive norms of our social relations, or at least those that constitute social relations of valuable kinds.

³ Rouse makes this clear in his example of nursing practices (2023: 137-8). Responding to the “constitutive concerns” of nursing, he says, “is neither a *de facto* regularity” in behaviour, nor a response to a principle “that perhaps no nurse actually satisfies,” but rather “a situated normative concern that arises both from how nurses generally do comport themselves and what concerns they or others hold themselves accountable for.” In a jurisprudential idiom: it is a condition of the validity of the norms of social relations that they are socially effective, so that, in Jürgen Habermas’s (1996: 30) words, they have “*de facto* validity as measured by average acceptance.”

For norms to be in practice, I said, people have to uphold them *as if* something was at stake in the kind of social relations those norms constitute. This does not imply that those who uphold the norms must believe the relevant social relations matter, though it is consistent with such belief. To see this, consider that, in general, to treat *a* as if it were *F* does not imply believing that *a* is *F*. If I invite a close friend over for Christmas lunch, I may, along with my whole family, treat my friend as if he were family, without any of us believing that he is in fact family.⁴ Does this suggest, conversely, that social relations can be in practice without anyone who upholds their norms believing that there is any worth in doing so? I think not, because I do not see why people would, in general, come to pattern their behaviour in such determinate ways unless at least some of them have an idea of the value in the social relations thereby constituted, and they exhibit the idea of this value in how it guides their own behaviour. Many people's behaviour could surely not become so determinately patterned as to amount to upholding norms unless there are some, who they emulate, who uphold the norms in light of whatever they believe is at stake in upholding them. But only the faithful vanguard need have an idea of the value of the social relations they practice. The rest need only practice them *as if* they really matter, even if they have no idea quite how.

Thus, for norms to be in practice, there needs to be a vanguard of practitioners who have an idea of the value of the social relations thereby constituted. Of course, for a given social relation, this idea may very well be misguided. Nevertheless, for the norms to be in practice, it is enough that the vanguard uphold them in light of some idea of the value at stake, misguided or not. They need only believe that the social relation at issue matters, no matter the truth of that belief. Whether or not norms (and so social relations) are in practice, whether they exist, is therefore a sociological fact. This sociological fact about norms and the social relations they constitute does not alone have a bearing on whether they are justified. To be justified, what the vanguard believe is at stake—which the rest treat as if the case—would have to be genuinely valuable, and also genuinely at stake in behaviour that upholds the norms in question.⁵ Where that is so, the norms instituted

⁴ I borrow this treatment of 'treating as if' from work in a quite different context by Jessica Pepp, Eliot Michaelson and Rachel Katharine Sterken (2019: 69-70).

⁵ According to Han van Wietmarschen (2021: 75), the ability to separate what justifies social norms from the practices in virtue of which they exist makes Cristina Bicchieri's (2006; 2017) influential view of social norms superior to that of Geoffrey Brennan, Lina Eriksson, Robert E.

by the practitioners will also be supported by reasons, ones that reflect values genuinely at stake in upholding the norms.⁶

With these preliminaries in place, let me explain how I think our security interest interacts with our social relations. To simplify, I will assume that social relations obtain between two parties. People who stand in some social relation are bound to one another by certain directed obligations and rights, its constitutive norms, amongst which will be norms that give them authority over one another's actions. Sometimes it happens that one party, the *respondent*, acts in a way that sets back the security interest of the other, the *claimant*. The respondent acts in such a way that the claimant now has to invest new efforts—efforts that may have been earmarked for elsewhere in life—if they are to keep their life on track, and continue to meet the demands of what matters. Given the finitude of the claimant, the respondent's action makes it likelier that the claimant's efforts will be spread too thin, with all the associated psychological ramifications. In this way, the respondent's action sets back the claimant's security interest.

Fairly clearly, then, it is in the claimant's security interest for it to be part and parcel of their social relations that they have a claim against the respondent for acting as they did. After all, if, in so acting, the respondent fails to do what they owe it to the claimant to do, the claimant effectively loses an investment of effort on the respondent's part, and with it the ability to sustain the benefits of that investment (instrumental or otherwise) without reallocating efforts of their own. But it is an investment that the claimant can recoup. In breaching their obligation to the claimant, the respondent incurs a reparative obligation, such that it falls to them to make amends.⁷ They do so by making up, as far as is now possible, for what they originally owed it to the claimant to do. In the ideal case, that would remove the need for the claimant to reallocate their own efforts, so that such incidents have no tendency to spread them too thin.

Goodin and Nicholas Southwood (2013). However, the Bicchieri view—even suitably modified to address van Wietmarschen's criticisms (2021: 77)—seems to make accountability accidental to the idea and existence of social norms, which is, I take it, implausible. The picture put forward here, inspired by that of Rouse (2023), does not suffer this shortcoming.

⁶ For a recent and congenial account of our power to institute social norms and social orders in the narrower jurisprudential context of our so-called 'constituent power' to institute, and thus legitimate, legal orders, see N. P. Adams (2024).

⁷ For more on reasons of repair, see the preamble to Part Two.

Less clearly, perhaps, if the claimant's security interest is protected in the above way, this cannot but set back the security interest of the respondent. It may sound improbable, unpalatable even, that the party responsible for setting back the other's security interest in the first place is also exposed to insecurity by their social relations. For it may seem that we are, as it were, making a victim of the aggressor. I believe this worry is misplaced. Not all conduct that sets back another's security interest can be sensibly described as an act of 'aggression.' As such, if we can point to how the arrangement might set back the security interest of respondents, I see no obvious ethical reason not to direct our concern to them as well. Such a setback is readily observed. We saw that if the social relation's constitutive norms furnish a claim against respondents, it falls to them to make amends. Repair is itself hard work, requiring an investment of efforts—efforts that may have been earmarked for elsewhere—that makes it likelier that the respondent will be spread too thin, and correspondingly vulnerable to inner turmoil.

I claim, therefore, that action governed by our social relations may set back security interests on either side of the equation, and that those setbacks are systematically linked. In the wake of an incident, the *ex post* burdens of effort are divided between claimants and respondents in a zero-sum fashion. If the respondent has no obligation not to act as they did, then the *ex post* burdens of effort will remain where they fell: in the claimant's lap. But if the respondent does have such an obligation, then it will fall to them to make amends, absorbing from the claimant as much of the *ex post* burdens of effort as possible. Any *ex post* burdens of effort that lie with the respondent are removed from the claimant, and vice versa. This is how the constitutive norms of our social relations distribute the security costs zero-sum between the parties, effecting what I call a *local* distribution of security.⁸ The local distribution of security in social relations of given kinds depends on how we set up their constitutive norms, that is, on which norms are in practice amongst the practitioners—the vanguard and the rest.

I have been focusing on norms set up to protect claimants against security costs for which the respondent is responsible, irrespective of whether the respondent incurs those costs intentionally, recklessly or negligently, i.e. regardless of whether they were at fault.⁹ This is a simplifying assumption. Fault seems to be an ingredient of the breach of at least

⁸ This terminology is adapted from Stephen Perry (1992: 461).

⁹ See Chapter One for discussion of this notion of 'fault.'

some obligations, such that, if the respondent is faultless then they cannot be in breach. As Gardner says (2005: 115), “Being at fault sometimes contributes constitutively, and not just instrumentally, to the wrongfulness of one’s actions. Some wrongs [...] are ‘fault-anticipating.’”¹⁰ Conversely, I will call an obligation ‘strict’ if as it is not fault-anticipating in Gardner’s sense.¹¹ Obligations to ϕ can be strict, requiring us to ϕ come what may, or they can be fault-anticipating, requiring us to ϕ except where our failure to ϕ would be faultless. I put this by saying that obligations can have the same matter whilst differing in their *form of liability*. So far, then, I have abstracted from variation in forms of liability, assuming that our social relations only ever furnish respondents with strict obligations owed to claimants.

On dropping this assumption, another way that our social relations can effect a local distribution of security comes into view.¹² It is one thing to ask if obligations should be set up to protect claimants against respondents acting in certain ways, and quite another, if they should, to ask if they should be strict or fault-anticipating. The decision as to the form of liability impacts the localised distribution of security no less than the more basic decision as to whether there should be an obligation at all. We have seen how deciding against an obligation systematically favours respondents. I claim that, even if we decide in favour of an obligation, its being fault-anticipating systematically favours respondents in much the same way. To see this, suppose that the claimant enjoys the protection of a strict obligation against the respondent’s having ϕ -ed. In that case, no matter whether the respondent was at fault in having ϕ -ed, their having ϕ -ed entails breach of obligation,

¹⁰ Fault-anticipating obligations may seem incoherent on the grounds that the notion of ‘fault’ applies only if we have already done wrong, and so cannot be amongst the wrong’s ingredients. In Gardner’s view (2005: 116), fault-anticipating wrongs are always parasitic on *something else* we were at fault in doing, in breaching the fault-anticipating obligation—perhaps our failure to satisfy a strict obligation, or even a weighty but non-obligatory reason. Gardner’s (2001: 134-43) argument for the parasitism of fault-anticipating obligations builds on the work of Tony Honore (1999: 29).

¹¹ I follow Stephen Smith (2014: 190) in this use of ‘strict.’ See Chapter Two for a defence of the coherence of strict obligations against what I call ‘moralism’ about obligation.

¹² The points made in this paragraph are inspired by Gardner’s (2014: 211) discussion of the distributive consequences of the choice of liability regime in private law contexts: “What the defendant gains from there being fault liability, the plaintiff loses; what the defendant loses from there being strict liability, the plaintiff gains.”

so that they now need to make amends. Now suppose the obligation is fault-anticipating. In that case, there is a further question to ask before we can determine breach, namely, was or was not the respondent at fault in having ϕ -ed? If they were not, then there is no breach—the obligation is, after all, fault-anticipating—and no need for repair on their part. This works to the detriment of the claimant’s security interest, since they must now reallocate their own efforts if they are to keep their life on track, making it likelier that they will be spread too thin. In other words, fault-anticipating obligations systematically favour respondents by shifting the security costs of faultless accidents towards claimants; strict obligations shift the same costs towards respondents.¹³ Needless to say, these are costs from the perspective of our security interest, an interest which has both material and psychological dimensions.

7.3. The Distinctively ‘Political’

Thus, for given kinds of social relations, necessarily: their constitutive norms distribute security locally, between parties to them, in a zero-sum fashion. I claim that the idea of what is at stake in those social relations—and so the justification of their norms—can therefore rightly include considerations as to the proper distribution of security between

¹³ Since the norms of our social relations are not legal norms, my distinction between fault-anticipating and strict obligations should not be assumed to neatly correspond with torts that are and are not subject to the negligence standard, i.e. to the tort law ‘duty of care.’ To see this, take the decision of the High Court of Australia in *Burnie Port Authority v General Jones Pty Ltd* ((1992-94) 179 CLR 520). In *Burnie*, contractors for Burnie Port Authority caused a fire that spread to neighbouring premises, ruining the plaintiff’s frozen vegetables. The High Court concluded that Burnie Port Authority was liable in negligence, despite the absence of personal fault on their part, due to a failure to meet their “non-delegable duty” to take positive action to ensure that care was taken, in this case, by their contractor. In Jenny Steele’s analysis (2022: 748), “Through its use of positive duties to ensure that care is taken, the *Burnie* case accepts that there is nothing outdated about a rule whose effect is to impose liability in the absence of personal fault”. As she sees it, the tort law duty of care is flexible enough to legally prohibit conduct that is quite faultless from an extra-legal perspective. So, though Burnie Port Authority was negligent in tort—i.e. in breach of its fault-anticipating *tort law* obligation—it was without personal fault, and so, *a fortiori*, not in breach of fault-anticipating obligation *outside* the law. To reiterate: I am interested in our social relations outside the law.

the parties. These considerations concern the *justice* of the local distribution of security in our social relations.

I follow Gardner (2000: 7) in thinking of ‘justice’ as a concern about allocation, that is, a concern for “who gets how much of what and why.” The ‘what’ that is to be allocated may or may not be scarce, and this determines ‘how much’ can be allocated. Where the ‘what’ is scarce, as it is with security between the parties, the concern with ‘how much’ should be allocated to whom is a distributive concern insofar as it takes that scarcity into consideration. The ‘who,’ in our case, is limited to the parties to whatever social relations are to be determined, considered in their role within those relations: landlord or tenant, parent or child, and so on. This is what makes the distribution being considered a local one. We want to know whether it is landlords or tenants, parents or children, who should be exposed to insecurity (and to what extent) by incidents that happen in the context of their relations.

In saying ‘why’ they should be exposed to whatever degree the norms dictate, we state putative grounds for the justice of the local distribution of security between them. Familiar such grounds would include, for example, that we should distribute security so as to maximise it on net, to give priority to the most insecure, to ensure enough of it for all, or to divide it up equally. These familiar grounds for the justice of local distributions admit of at least two interpretations, depending on how we interpret their scope. They may enjoin us to set up our social relations so as to maximise, prioritise, sufficientise, or equalise security either within the narrower scope of parties to those relations, or within the wider scope of society at large. To take maximisation as an example: it is one thing to set up social relations so as to minimise the net extent to which conduct within their sphere will spread the parties too thin, and quite another to minimise the net extent to which conduct within their sphere will spread people in wider society too thin.

I want to suggest that, however interpreted, these familiar grounds for the justice in local distributions of security are in a certain sense *apolitical*. First of all, the grounds of allocation make no reference to who, in society, stands to benefit from the allocation, that is, in which group’s security interest it would be for social relations to be set up in a certain way. I claim that for considerations of local security-distributive justice to count as distinctively political, they need to answer the question of ‘why’ with reference to a group in whose security interest it would be for the social relations at issue to be set up

in the manner proposed.¹⁴ Of course, the mere mention of any old beneficiary group in our considerations will not on its own tend to bear on the justice of the proposed norms. To do that, it needs to come bundled with a judgment about how benefitting that group might redress societal imbalances between different groups and their security interests. For instance, setting up norms on the grounds that they advance the security interest of groups who are already relatively well-protected in the current social order only serves to entrench existing imbalances, and the resulting local distribution of security does not strike us as plausibly grounded in justice.

To make judgments like this, however, we need to look beyond the social relations being determined, to where they sit within the wider constellation of social relations, to their place in the social order. This is the case even if we interpret the group benefitted narrowly, that is, if we take the group to figure as a relatum in the social relations at issue. Take the question of whether the landlord-tenant relation should protect tenants against landlords failing to make housing repairs. Even if we answer ‘yes’ on the narrow basis that it is in tenants’ security interest, and that the social position of tenants renders them insecure relative to landlords, we need to back this up with some judgment about how people who are tenants and people who are landlords are generally faring, security-wise, in the social relations they typically find themselves in. To see this, imagine a social order in which those who are landlords are generally poorly protected against insecurity, and those who are tenants are well-protected. In that social order, the fact that further protections for tenants would advance that group’s security interest might not strike us as making those protections just, in the way they distribute security locally.

This speaks to another mark of the apolitical character of the more familiar grounds of local security-distributive justice. Considerations to the effect that setting up certain norms in a certain way will maximise, prioritise, sufficientise or equalise security offer grounds of distributive justice that can be stated as rules or principles of justice. As such,

¹⁴ This chimes with how Raymond Geuss develops Lenin’s definition of ‘the political.’ Geuss (2008: 23) thinks that “Lenin defines politics with characteristic clarity and pithiness when he says that it is concerned with the question that keeps recurring in our political life: ‘Who whom?’ (кто кого).” But although Lenin’s formula, he says (2008: 25), is “basically correct,” it is “perhaps too dense and needs to be developed or extended.” He proposes (ibid.) to extend it in at least the following way: “the formula should read not merely ‘Who whom?’ but, rather, ‘Who <does> what to whom for whose benefit?’ with four distinct variables to be filled in, i.e., (1) Who?, (2) What?, (3) To whom?, (4) For whose benefit?”

I believe, they reflect an ‘abstract’ form of justice, in the sense that they offer grounds of distribution that are applicable, not just to *this* social order—comprised of individuals standing in *these* social relations—but to *any* social order. By contrast, properly political grounds of distribution, those that appeal how certain arrangements stand to benefit certain groups, are ‘concrete.’ Whether a given group stands to benefit from the norms proposed, and whether benefitting that group in that way would serve to redress existing imbalances in the social order, is contingent on how that social order is constituted, and where the social relations at issue fit into the wider constellation. What works for certain relations in one society may not work in another. Correspondingly, the kind of judgment required to assess the local security-distributive justice of given norms does not involve measuring them against an abstract rule or principle, one applicable to any social order, but rather an attunement to how they concretely interact with the particular social order in which they figure.¹⁵

I recognise how this appeal to the ineffable particularity of social reality may strike some as hopelessly obscure. Surely, we may think, a kind of ‘justice’ whose content is not expressible as a rule or principle is no justice at all. I would reply, with Gardner, that this worry is symptomatic of an unduly legalistic conception of justice. Drawing some lessons from Aristotle’s remarks on the virtue of ‘equity,’¹⁶ Gardner says (2000: 17),

some just rulings are not governed by nor capable of being elevated to any sound rule of justice. They are based on a weighing of allocative considerations in their raw, unruly form. Solomon’s justice was justice *ad hoc*, and none the worse for that. Likewise the justice of a modern-day arbitrator who, unlike a judge in a court of law, looks at the merits of the case before him without being bound to

¹⁵ I regard this view as continuous with Lorna Finlayson’s (2020: 121) preferred methodology in political philosophy: “The way in which reflection on the real world should direct our theory won’t always be reducible to a formal argument [...]. It may more often be a matter of the making of judgments that are barely expressible in a discursive form at all.” She is aware that this idea may leave some readers “rolling their eyes or mentally drifting off,” but she stands by it, if only because, as she says, “I don’t think it’s possible to say anything much more precise” (ibid.). I agree, though I will suggest that the idea can be made more palatable by situating it in a less heterodox tradition.

¹⁶ NE 1137a32-1138a2.

explain how his decision on this case has been or would be generalized to any other decision he might make on any other case.

For Gardner, then, the justice of an arbitrator may not be expressible in an abstract rule, but it is no less a kind of justice for it. By the same token, I believe that good political judgment is responsive to social reality as it confronts us concretely, making the content of the resultant considerations of justice inaccessible to the application of abstract rules or principles. The concrete local security-distributive justice of norms set up sensitively to those considerations cannot be stated accurately in any abstract rule or principle. To try to do so, Gardner suggests (2000: 18), is something on which “legal systems, being systems of rules, will tend to insist [...] a tendency which, for the sake of justice itself, sometimes needs to have its wings clipped.” I believe that to avoid a similarly legalistic view of local security-distributive justice, we must accept the normativity of distinctively political considerations, which reflect judgments about what justice concretely requires given the relative security of different groups in the existing social order.

Let us walk through an example of the distinctively political considerations of local security-distributive justice I have in mind. Recall the concrete situation at the Dolphin House estate. The tenants there were at the sharp end of a decline in standards of housing administered by Dublin City Council (DCC), and the Irish government more broadly. Suppose the dreadful conditions at Dolphin House set back the security interest of the tenants.¹⁷ Does DCC let the tenants down in letting its housing stock get so decrepit, or should the tenants just get on with it? If the norms of landlord-tenant relations provide that DCC do indeed fail in their housing obligations, this exposes DCC to insecurity. It falls on them, as far as possible, to make reparative efforts, relieving tenants of the effort of resolving their housing situation themselves. But if the norms say instead that the tenants are asking too much, then they will be the ones exposed to insecurity, since they, and not DCC, will be the ones to reallocate their efforts to resolve the housing situation. By now this will be familiar: the norms of the landlord-tenant relation distribute the risk of being spread too thin by housing disrepair, the security costs, in a zero-sum fashion.

¹⁷ According to a complaint filed with the European Committee of Social Rights, the tenants experienced “sewage invasions” of “grey and black wastewater,” resulting in “dampness and fungal contamination” by a “known human pathogen.” See case study 1, section III B.1.3.1 of Complaint No. 110/2014 *International Federation for Human Rights (FIDH) v. Ireland*.

Now consider that tenants may also be parents of dependent children. Suppose that the norms of the landlord-tenant relation are such that the tenants were asking too much of DCC, so that it falls to them to resolve the housing situation. But with bills to pay and mouths to feed, it would be no wonder if their efforts are spread too thin to meet their obligation to safely house their kids. Though they are claimants in their relations to DCC, they are respondents as parents to their kids. We may well want the whole arrangement to work to the benefit of children, safeguarding their security interest. If so, it will seem undesirable that we set up our parent-child relations such that children are exposed to the security costs that attend the circumstance of unsafe housing. The corresponding local insecurity of the tenants *qua* parents is clearly relevant to who insecurity should be locally distributed to in their relation with DCC. After all, if tenants must, *qua* parents, absorb the security costs of unsafe housing from their kids, it may seem unjust to have them also absorb the costs in their relations with landlords, especially if landlords in fact enjoy relative security as a group. When we think in this way—about how security flows through the social order, letting that guide how we set up its constituent relations—we are bringing distinctively political considerations of justice to bear on how we distribute security locally.

Considerations of recognisably the same shape could, of course, foreground other beneficiaries than children. Suppose that landlords owe no housing obligation to tenants, so that the tenants absorb the security costs of shoddy housing. To whose benefit would this be? On the narrowest interpretation, landlords like DCC stand to benefit. There may be all sorts of pressures on those who are landlords beyond those they face as landlords, and locally shifting insecurity towards tenants may leave landlords less exposed to being spread too thin in these other areas. In the case of DCC this would include various areas in its remit *qua* local authority. But to widen the interpretation, note that there are others, besides DCC, whose security interest is served by the tenants' exposure in the landlord-tenant relation. Plausibly, central government is obligated to help fund the efforts of local authorities sufficient to meet demands on the ground, under the norms of their relations. If local authorities' efforts are spread too thin, central government must therefore work to help fund them adequately. This in turn requires an effort from central government—efforts that may have been earmarked for elsewhere—which, of course, tends to spread central government too thin. Thus, should we wish to foreground central government's security interest, we might make peace with the insecurity of tenants, if

we perceive that the corresponding security of local authorities *qua* landlords helps to prevent central government from being spread too thin.

The point of the example is not to recommend a particular stance on this fragment of Irish social order, or to set up the landlord-tenant relation in a particular way. Nor is it to suggest that the two stances we considered are the only possible or intelligible ones available. Rather, my point is that, whatever beneficiary we foreground, and however we see the interconnections between different relations in the social order, we are bringing distinctively political considerations of justice to bear on the norms of landlord-tenant relations, insofar as they distribute security zero-sum between the parties. What makes them distinctively political is that they appeal to how benefitting a certain group might serve to redress societal imbalances, a judgment based on an attunement to social reality in all its concreteness.

In closing the section, I want to address one final worry. Even if our social relations necessarily effect local distributions of security, we may wonder if we should actually let such distinctively political considerations guide how we practice them—or whether the vanguard should, at least. Does it not make us guilty of unduly *politicising* what matters, introducing into our practices something properly extrinsic to their value? I am not sure we should always think of such political considerations as extrinsic in this way. It really seems to depend on the concrete circumstances. Consider the teacher-student relation in higher education today. Perhaps once upon a time, when there was less pressure on the university system—from the democratisation, commodification and marketisation of tertiary education—what it meant to be a good teacher could float completely free of attunement to political considerations. Teachers from that era might look upon today's teachers as degraded, distracted in their practices by political concerns. But is it so clear that teachers today should refer themselves to standards fit for a bygone university system, swinging completely free of the circumstances of the present? I do not think so. This is not to deny that excess is possible in the attunement of today's teachers to political concerns; what counts as such gets played out in the norms that today's teachers uphold. But it surely could be the case that good teacherly practice, under today's pressures, cannot be achieved without sensitivity to such concerns. To think otherwise strikes me as a sort of nostalgia about teacher-student relations that makes their special value the preserve of an elite, namely, those who can afford to teach or learn unfettered by today's pressures.

7.4. Politics and Forms of Liability

In section 7.2 I said that how our social relations distribute security locally depends in two ways on how we set up their constitutive norms: they may furnish respondents with obligations not to act in certain ways that incur security costs on claimants, and, if they do, the form of liability of those obligations may be fault-anticipating or strict. Let us return to the circumstances of Dolphin House to see how decisions around the form of liability may also be sensitive to the distinctively political considerations of localised security-distributive justice we just discussed.

It is hard to tell the story of how conditions in estates like Dolphin House came to be so awful without some reference to the failure of local authority housing regeneration projects, and the experiment in public procurement policy that formed the backdrop. One human rights organisation recounts that “from 2001 onwards, Public Private Partnerships (PPPs) were promoted as the principal mechanism to create regeneration,” freeing up funds by “land transfer arrangement between local authorities and private developers, alongside private investment.”¹⁸ The official government rationale for the use of PPPs is to distribute risk more attractively for private sector investors, so as to harness their dynamism and achieve value for money for the public purse. A perhaps less well-advertised rationale is that it allows for public services and infrastructure projects to be financed without showing up on the balance sheet, thereby serving the government of the day’s interest in claiming fiscal responsibility to its electorate.¹⁹ So when, following the 2008 financial crisis, “all of the PPP projects collapsed,”²⁰ we might take the meaning of PPPs for relations between central and local authorities to be relevant to what DCC should owe its tenants.

¹⁸ *FIDH v. Ireland* (2014), section III C.

¹⁹ At least in the UK case, the cross-party Treasury Select Committee concludes of the ‘Private Finance Initiative’ (PFI), a PPP policy, that “The incentive for government departments to use PFI to leverage up their budgets, and to some extent for the Treasury to use PFI to conceal debt, has resulted in neglecting the long term value for money implications.” See page 35 in the 2011 report, available here: <https://publications.parliament.uk/pa/cm201012/cmselect/cmtreasy/1146/1146.pdf> (accessed March 20th 2025).

²⁰ *FIDH v. Ireland* (2014), section III C.

For example, we could take the Irish government's use of PPPs as signalling that it no longer holds itself obligated to fund its local authorities, or at least not in a way that requires provision of adequate taxation-back funds to meet needs on the ground. After all, this new procurement policy in effect leaves local authorities to fend for themselves, as DCC did by exposing itself to global financial markets. The relations between central government and DCC is thus not one whose norms protect DCC from the security costs of inadequate funding, and so, in that event, they must reallocate their efforts if they are to meet (for instance) the housing needs of its tenants, which they did using PPPs. This may seem relevant to how we should set up protections for the tenants. Even if we think that DCC's inaction sets back the security interests of its tenants, we may also perceive that the need for private investment, and so exposure to its collapse, is hardly DCC's fault. We may, in turn, consider this relevant to whether DCC should be exposed to yet more insecurity, this time vis-à-vis its tenants. For instance, we may think this would be fine if DCC were at fault, but that it is too much to burden them with the effort of repair if they are faultless. In thinking like this, we are taking how security is locally distributed by the norms governing intra-governmental relations to bear on the form of liability encoded in norms governing the landlord-tenant relation.

We could, of course, arrive at a different conclusion by considerations of much the same shape. If, as suggested, we set up DCC's housing obligation as fault-anticipating, this works systematically to the disadvantage of the tenants, leaving them exposed to the burden of reallocating their own efforts in the event of shoddy housing. This will tend to spread them too thin, for example, to meet obligations they may have *qua* parents to safely house their kids. Or it will do, at least, if those obligations are strict. For if, instead, they are fault-anticipating, then the failure to safely house their kids will presumably not amount to breach, since the tenants seem no more at fault than does DCC. Plausibly, however, the kids cannot be expected to shoulder the security costs—or not, at least, without radically revising our understanding of parent-child relations. If we take that as a fixed point, the parental obligation must be made strict: the parents, not the kids, should shoulder the security costs of faultless accidents. Once again, this seems relevant to the form of liability we recognise in DCC's housing obligation to its tenants. For if we make the obligation strict, then DCC, not the tenants, will be the ones to absorb the security costs exacted by the sorry state of Dolphin House, thereby relieving the tenants of reparative efforts they would otherwise owe their kids for failing to safely house them.

In each case, we take a stance on who benefits from the whole arrangement. Where DCC's housing obligation is set up as fault-anticipating, it is central government—more than the local authority—who really appears to be the main beneficiary. But the flipside is that this burdens the tenants with fixing their housing situation, which tends to spread their efforts too thin, for instance, to safely house their kids. Such an arrangement, then, seems to work to the benefit of Irish central government and to the detriment, ultimately, of those we like to think should not be expected to fend for themselves. By furnishing landlords with fault-anticipating rather than strict housing obligations to their tenants, we are, in effect, siding with those who have the privilege to govern over and above the most vulnerable in society. We may, in a paradigmatically political reflex, recoil from the way this arrangement balances the security interests of different groups in the social order. If we wish to prioritise children's security interests before anything else, we may be moved instead to practice a strict form of liability in landlords' housing obligations to their tenants.

I do not suggest that these are the only distinctively political considerations we can formulate regarding the forms of liability encoded in the norms discussed. There are, no doubt, many other stances that we could take. But the two we walked through represent perhaps the paths of least resistance, at least if we hold fixed the strictness of parents' obligations to safely house their kids. We held this fixed because it seemed necessary to preserve deeply held convictions about the value of parent-child relations, convictions that are in this way beyond political contestation. Everyone agrees that part of what it is to be a parent is to shoulder an asymmetrical burden of insecurity in their relations with their infant children, at least to sort out matters such as housing. In closing, and in line with our political theme, I wish to explore another way we might practice the forms of liability in the norms of landlord-tenant relations, holding fixed our convictions about the parent-child relation.

There is, I claim, a crude type of 'liberalism' that regards the very concept of a strict obligation as anathema, on grounds of moral metaphysics.²¹ It begins with the claim that we each have equal dignity in virtue of being rational agents, free to exercise our powers of rational agency consistently with everyone's freedom to do the same. Our obligations

²¹ This is one way of arriving at the general view I call 'moralism' about obligation, which I argue against in Chapter Two.

get their moral force from this dignity, and so their rationale is always that they in some way protect the free exercise of our powers of rational agency. We are, however, material beings in a material world, limited by the use of our bodies to pursue our ends, on the one hand, and by finite space and other scarce resources on the other. As such, the free exercise of our rational agency may in principle pose a risk to others' freedom to exercise theirs, and with it, their being treated with the respect their dignity merits. Thus, on this view, the interactions of private persons are governed by moral interpersonal obligations, responding to breaches of which is the substance of the right and the just. For this type of liberal, parent-child interactions also belong to the domain governed by interpersonal obligations with this general moral rationale.²²

The problem with strict obligations, on this view, is that they are fundamentally at odds with the rationale of protecting the free exercise of our powers of rational agency. As Ernest Weinrib (1995: 181) puts it,

Strict liability [...] implies that the very production of external effects—an indispensable part of agency—can itself be a violation of the equality of agents. [...] Effects are merely the fruition of activity. To ascribe liability to an action, regardless of culpability, for whatever harmful effects it has had simply because they *are* its effects, is to hold the agent liable for being active.²³

The thought here, roughly, is that for material beings like ourselves—embodiment and all—the exercise of our powers of rational agency is, *as such*, an efficacious material

²² Even Kant's *Metaphysics of Morals* (1996), in the Doctrine of Right, includes, alongside fundamental private rights to property and contract, what he calls "*rights to persons akin to rights to things*" (MM 6: 282), which govern various relations of dependency characteristic of the domestic sphere, including parent-child relations. Helga Varden (2012: 336 -7) calls these 'status rights,' noting that, "For Kant, the core philosophical challenge is how to conceive of such standing in unified or shared private lives between equals, as well as between equals and unequals, in a way reconcilable with what he calls everyone's 'innate right to freedom.'"

²³ We may worry that discussing Weinrib is not to the point here, as my express interest is in the norms of social relations considered extra-legally, whereas Weinrib is concerned with private law relations. But Weinrib subscribes to the characteristically Kantian view that morality is in an important sense incomplete without law, such that legal norms (or at least the good ones) are moral norms, or, as he (1995: 110) puts it, "where practical reason formulates ethical duties, juridical ones have already taken hold."

process. It *just is*, at least in part, to cause effects in the world. But strict liability descends upon us irrespective of fault, that is, simply due to the worldly efficacy of our powers of rational agency, whether or not we exercised them aright rationally speaking.²⁴ Since the breach of strict obligations is a ground of strict liability, it would appear, then, that strict obligations proscribe exercises of our powers of rational agency purely on the basis of a worldly efficacy that is native to them as such. So, the thought goes, far from protecting the free exercise of our powers of rational agency, strict obligations in fact prohibit such exercises for being what they are. The crude sort of liberalism I have in mind is animated by a line of thought like Weinrib's, or something close to it, about the moral foundations of our social relations.²⁵

The crude liberal, like the rest of us, will want to preserve our sense that children living in conditions like those in Dolphin House should not be exposed to the associated security costs in their relations with their parents. For parent-child relations to do justice to our prevailing understanding of their value, it has to be that the conditions at Dolphin House put the parents in breach of an obligation to safely house their kids. But since the crude liberal denies the existence of strict obligations, they must take this obligation to be fault-anticipating. Thus, the only way they can conceive of the parents as in breach is for them, in fact, to have been at fault in failing to safely house their kids. But in the concrete circumstances of the case—the global financial crisis, investor withdrawal, and subsequent collapse of PPPs—this view looks extremely difficult to sustain. Indeed to insist on such an assessment may strike us as profoundly lacking in political judgment, an appearance that is only compounded by its alleged basis in some special insight into the nature of morality.²⁶

Recognising this, perhaps the crude liberal can try to apply a similarly demanding threshold for faultlessness to DCC's housing obligations, so that they absorb the security

²⁴ See Chapter One for discussion of the notion of 'fault' along these lines.

²⁵ See Peter Cane (1996: 485-7) for several lines of puzzlement on the cogency of Weinrib's argument and whether, even if cogent, its conclusion coheres with his overall picture of private law. I remain agnostic about whether Weinrib's argument rules out Kantian justifications for 'use-based' strict obligations in private law (e.g. property, trespass), or, as Arthur Ripstein (2007: 18) suggests, it does not.

²⁶ I discussed this sort of fault-finding in Chapter Two as one of the excesses of what I there called 'moralism' about obligation.

costs in practice, thus relieving the parents of the risk of breach themselves. The problem with such a view is that, in the circumstances, it is no less implausible that DCC was at fault than that the parents are. So, the basic problem with the view remains. In order to deliver what strikes us as the right verdict on the proper distribution of security between the parties to relations of dependence, crude liberalism strains our convictions about the lengths to which people and institutions can be expected to go—how much grit or savvy they must show—before they are absolved of the charge of characterological vice. Having deprived themselves of one of the main ways of establishing security-distributive justice in social relations, the crude liberal is forced to do politics by the only means they have left: to moralise *ad hoc* about individuals and their deficit of virtue.

7.5. Concluding Remarks

I have sought to show that the constitutive norms of social relations, including the forms of liability they encode, distribute security between the parties to them in a zero-sum manner. This makes how we practice them sensitive to considerations of justice in the local distribution of security between the parties, considerations that have a distinctively political character when they reflect a judgment about who benefits, security-wise, from setting up the norms in a certain way, given how the social order is concretely arranged. There is, then, a distinctively political dimension to the norms we practice in our social relations, including their forms of liability. They exhibit this dimension simply because they jostle for attention in the lives we lead, lives that, in their finitude, are necessarily at risk of being spread too thin—with all the psychological ramifications that entails.

It is this last point, I suggest, that connects this chapter's discussion with the wider project of Part Three, which is to vindicate the slogan: 'Tragedy has a politics.' This is because, as I have argued, whether our social relations furnish us with strict obligations is subject to considerations of a distinctively political shape. When our social relations strictly obligate us to do certain things, we are exposed to relative insecurity: it falls to us to make up for our failure, even if we failed through no fault of our own. What we have so far left unrepaired is an apt object of agent-regret. And since repair itself requires an investment of effort—efforts that may have been earmarked to meet demands we face elsewhere in our lives—the need to repair will tend to spread us too thin, which will, by

the same route, make us apt feel yet further agent-regret. Strict obligations in our social relations thus expose us to tragedy in many ways. In failing to live up to them, we come to be, through no fault of our own, apt to feel the painful self-directed sentiments that characterise the phenomenology of tragic situations.

I argued that whether we have strict obligations in the first place is in part a political affair. For it can be influenced by the considerations we bring to bear on how we practice the constitutive norms of our social relations, considerations that may be distinctively political in that they reflect a judgment about how the security interests of different groups are balanced in the existing social order, and how imbalances may be redressed in favour of certain groups by practising the norms in a certain way. Such considerations, I argued, are 'concrete' in the following sense: had social relations in the wider social order been organised differently, such that the same relations (with the same constitutive norms) worked primarily to serve some other beneficiary, the relevant considerations would have been different, potentially supporting norms distributing tragedy elsewhere in society. Tragedy has a politics because our social relations necessarily distribute security between the parties to them, parties who in turn stand in a wider constellation of such relations. Our social relations distribute security *at all* due to our worldly condition: though many things may matter to us, we only have so much effort to give.

Bibliography

- Adams, Nate. P. (2021). In defense of exclusionary reasons. *Philosophical Studies*, 178(1), 235-253.
- Adams, Nate. P. (2024). Constituent power-with. *Philosophy & Public Affairs*, 52(3), 289-326.
- Alexander, Larry. (1990). Law and exclusionary reasons. *Philosophical Topics*, 18(1), 5-22.
- American Law Institute. (1985). *Model Penal Code: Official Draft and Explanatory Notes*. The American Law Institute.
- Annas, Julia. (1995). *The Morality of Happiness* (2nd ed.). Oxford University Press.
- Arendt, Hannah. (2022). *Eichmann in Jerusalem*. Penguin Random House UK.
- Aristotle. (2014). Nicomachean ethics. In Jonathan Barnes and Anthony Kenny (Eds.), *Aristotle's Ethics* (207-371). Princeton University Press.
- Arneson, Richard. (2011). Luck egalitarianism—a primer. In Carl Knight and Zofia Stemplowska (Eds.), *Responsibility and Distributive Justice* (24-50). Oxford University Press.
- Arpaly, Nomy. (2002). Moral worth. *The Journal of Philosophy*, 99(5), 223-245.
- Austin, John. L. (1957). A plea for excuses. *Proceedings of the Aristotelian Society*, 57(1), 1-30.
- Baier, Kurt. (1966). Moral obligation. *American Philosophical Quarterly*, 3(3), 210-226.
- Baron, Marcia. (1988). Remorse and agent-regret. *Midwest Studies in Philosophy*, 13(1), 259-281.

- Bentham, Jeremy. (2007). *An Introduction to the Principles of Morals and Legislation*. Dover Publications.
- Betzler, Monika. (2015). Review of *The View From Here: On Affirmation, Attachment, and the Limits of Regret*, by R. J. Wallace. *Ethics*, 125(2), 614-621.
- Bicchieri, Cristina. (2006). *The Grammar of Society*. Oxford University Press.
- Bicchieri, Cristina. (2017). *Norms in the Wild*. Oxford University Press.
- Bittner, Rüdiger. (1992). Is it reasonable to regret things one did? *The Journal of Philosophy*, 89(5), 262-273.
- Boonin, David. (2008). *The Problem of Punishment*. Cambridge University Press.
- Brandt, Richard. (1958). Blameworthiness and obligation. In Abraham I. Melden (Ed.), *Essays in Moral Philosophy* (3-39). University of Washington Press.
- Bratman, Michael. E. (1999). *Faces of Intention*. Cambridge University Press.
- Brennan, Geoffrey, Erikson, Lina, Goodin, Robert. E. and Southwood, Nicholas. (2013). *Explaining Norms*. Oxford University Press.
- Callard, Agnes. (2017). The reason to be angry forever. In Myisha Cherry and Owen Flanagan (Eds.), *The Moral Psychology of Anger* (123-137). Rowman & Littlefield.
- Cane, Peter. (1996). Corrective justice and correlativity in private law. *Oxford Journal of Legal Studies*, 16(3), 471-488.
- Chang, Ruth. (2004). Can desires provide reasons for action? In R. Jay Wallace, Philip Pettit, Samuel Scheffler and Michael Smith (Eds.), *Reason and Value* (56-90). Oxford University Press.
- Cockcroft, George. (1971). *The Dice Man*. Talmy Franklin.
- Cogan, Morton. (1974). *Imperialism and Religion: Assyria, Judah and Israel in the Eight and Seventh Centuries B.C.E.* Scholars Press.
- Cook, Karen. S., Hardin, Russell and Levi, Margaret. (2005). *Cooperation Without Trust?* Russell Sage.
- Cooter, Robert and Eisenberg, Melvin. A. (1985). Damages for breach of contract. *California Law Review*, 73(5), 1434-1481.
- Dan-Cohen, Meir. (2002). *Harmful Thoughts: Essays on Law, Self, and Morality*. Princeton University Press.
- Dan-Cohen, Meir. (2008). Luck and identity. *Theoretical Inquiries in Law*, 9(1), 1-22.
- Dancy, Jonathan. (2004). *Ethics Without Principles*. Oxford University Press.
- D'Arms, Justin and Jacobson, Daniel. (2000). The moralistic fallacy: on the 'appropriateness' of the emotions. *Philosophy and Phenomenological Research*, 61(1), 65-90.
- Darwall, Stephen. (2006). *The Second-Person Standpoint*. Harvard University Press.
- Darwall, Stephen. (2010). Authority and reasons: exclusionary and second-personal. *Ethics*, 120(2), 257-278.

- Davidson, Donald. (1971). Agency. In Robert Binkley, Richard Brunaugh and Ausonia Marras (Eds.), *Agent, Action, and Reason* (3-25). University of Toronto Press.
- Dubber, Markus. (2015). *An Introduction to the Model Penal Code* (2nd ed.). Oxford University Press.
- Duff, Antony. (2007). *Answering for Crime*. Hart Publishing.
- Dworkin, Ronald. (2002). *Sovereign Virtue* (1st ed., 4th reprint). Harvard University Press.
- Earman, John, Roberts, John and Smith, Sheldon. (2002). *Ceteris paribus* lost. *Erkenntnis*, 57(3), 281-301.
- Elliot, Andrew. J. and Niesta, Daniela. (2009). Goals in the context of the hierarchical model of approach-avoidance motivation. In Gordon B. Moskowitz and Heidi Grant (Eds.), *The Psychology of Goals* (56-76). The Guildford Press.
- Elshtain, Jean. B. (1995). Feminism, family, and community. In Penny A. Weiss and Marilyn Friedman (Eds.), *Feminism and Community* (259-272). Temple University Press.
- Engels, Friedrich. (1975). *The Origin of the Family, Private Property and the State*. Trans. Alec West, Ed. Eleanor B. Leacock. International Publishers.
- Enoch, David and Marmor, Andrei. (2007). The case against moral luck. *Law and Philosophy*, 26(4), 405-436.
- Fettweis, Christopher. J. (2017). Unipolarity, hegemony, and the New Peace. *Security Studies*, 26(3), 423-451.
- Finlayson, Lorna. (2020). If this isn't racism, what is? The politics of the philosophy of immigration. *Aristotelian Society Supplementary Volume*, 94(1), 115-139.
- Fischer, John. M. and Ravizza, Mark. (1998). *Responsibility and Control* (1st ed.). Cambridge University Press.
- Fraser, Nancy. (2013). *Fortunes of Feminism: From State-Managed Capitalism to Neoliberal Crisis*. Verso.
- Fricker, Miranda. (2016). What is the point of blame? A paradigm based explanation. *Noûs*, 50(1), 165-183.
- Fried, Charles. (2007). The convergence of contract and promise. *Harvard Law Review*, 120(3), 1-9.
- Fried, Charles. (2015). *Contract as Promise* (2nd ed.). Oxford University Press.
- Friedrich, Daniel and Southwood, Nicholas. (2011). Promises and trust. In Hanoch Sheinman (Ed.), *Promises and Agreements* (277-294). Oxford University Press.
- Fuller, Lon. L. (1969). Human interaction and the law. *American Journal of Jurisprudence*, 14(1), 1-36.
- Fuller, Lon. L. (2002). The role of contract in the ordering processes of society generally. In Kenneth I. Winston (Ed.), *The Principles of Social Order* (187-206). Hart.

- Gardner, John. (2000). The virtue of justice and the character of law. *Current Legal Problems*, 53(1), 1-30.
- Gardner, John. (2001). Obligations and outcomes in the law of torts. In Peter Cane and John Gardner (Eds.), *Relating to Responsibility* (111-144). Hart Publishing.
- Gardner, John. (2002). Reasons for teamwork. *Legal Theory*, 8(4), 495-509.
- Gardner, John. (2003). The mark of responsibility. *Oxford Journal of Legal Studies*, 23(2), 157-171.
- Gardner, John. (2005). Wrongs and faults. *The Review of Metaphysics*, 59(1), 95-132.
- Gardner, John. (2007a). Justifications and reasons. In John Gardner (Ed.), *Offences and Defences* (91-120). Oxford University Press.
- Gardner, John. (2007b). The gist of excuses. In John Gardner (Ed.), *Offences and Defences* (121-139). Oxford University Press.
- Gardner, John. (2014). What is tort law for? Part 2. The place of distributive justice. In John Oberdiek (Ed.), *Philosophical Foundations of Tort Law* (335-353). Oxford University Press.
- Gardner, John. (2018a). *From Personal Life to Private Law*. Oxford University Press.
- Gardner, John. (2018b). The contractualisation of labour law. In Hugh Collins, Gillian Lester and Virginia Mantouvalou (Eds.), *Philosophical Foundations of Labour Law* (33-47). Oxford University Press.
- Gardner, John. (2019). As inconclusive as ever. *Jerusalem Review of Legal Studies*, 19(1), 204-224.
- Gardner, John and Jung, Heike. (1991). Making sense of mens rea: Antony Duff's account. *Oxford Journal of Legal Studies*, 11(4), 559-588.
- Gardner, John and Macklem, Timothy. (2004). Reasons. In Jules L. Coleman, Kenneth Einar Himma and Scott J. Shapiro (Eds.), *The Oxford Handbook of Jurisprudence and Philosophy of Law* (440-475). Oxford University Press.
- Gaspa, Salvatore. (2018). *Textiles in the Neo-Assyrian Empire*. De Gruyter.
- Geuss, Raymond. (2008). *Philosophy and Real Politics*. Princeton University Press.
- Goldman, David. (2014). Modification of the reactive attitudes. *Pacific Philosophical Quarterly*, 95(1), 1-22.
- Grant, Heidi and Gelety, Laura. (2009). Goal content theories: why differences in *what* we are striving for matter. In Gordon B. Moskowitz and Heidi Grant (Eds.), *The Psychology of Goals* (77-97). The Guildford Press.
- Greco, J. (1995). A second paradox concerning responsibility and luck. *Metaphilosophy*, 26(1-2), 81-96.
- Greenberg, Alexander. (2024). Awareness and the recklessness/negligence distinction. *Criminal Law and Philosophy*, 18(2), 351-367.
- Gur, Noam. (2018). *Legal Directives and Practical Reasons*. Oxford University Press.

- Habermas, Jürgen. (1996). *Between Facts and Norms* (1st ed., reprint). Trans. William Rehg. The MIT Press.
- Han, Byong-Chul. (2023). *Absence*. Trans. Daniel Steuer. Polity Press.
- Hardin, Russell. (2002). *Trust and Trustworthiness*. Russell Sage.
- Hart, Herbert. L. A. (2008). *Punishment and Responsibility* (2nd ed.). Oxford University Press.
- Hartman, Robert. J. (2016). Against luck-free moral responsibility. *Philosophical Studies*, 173(10), 2845-2865.
- Hartman, Robert. J. (2017). *In Defense of Moral Luck*. Routledge.
- Hartman, Robert. J. (2019). Moral luck and the unfairness of morality. *Philosophical Studies*, 176(12), 3179-3197.
- Hayes, Lydia. J. B. (2017). *Stories of Care: A Labour of Law*. Palgrave.
- Hearne, Rory. (2014). Actually existing neoliberalism: public-private partnerships in public service and infrastructure provision in Ireland. In Andrew MacLaran and Sinéad Kelly (Eds.), *Neoliberal Urban Policy and the Transformation of the City* (157-173). Palgrave Macmillan.
- Hieronymi, Pamela. (2020). *Freedom, Resentment, and the Metaphysics of Morals*. Princeton University Press.
- Hirschman, Albert. O. (1986). *Rival Views of Market Society and Other Recent Essays*. Viking.
- Hirschman, Albert. O. (1997). *The Passions and the Interests* (2nd ed.). Princeton University Press.
- Holloway, Steven. W. (2002). *Aššur is King! Aššur is King!* Brill.
- Holton, Richard. (1994). Deciding to trust, coming to believe. *Australasian Journal of Philosophy*, 72(1), 63-76.
- Honoré, Tony. (1999). Responsibility and luck. The moral basis of strict liability. In Tony Honoré (Ed.), *Responsibility and Fault* (14-40). Hart Publishing.
- Howard, Christopher. (2022). Forever fitting feelings. *Philosophy and Phenomenological Research*, 107(1), 80-98.
- Howatt, Bill. (2022). *No Regrets: How to Live Today for Tomorrow's Emotional Well-Being*. Friesen Press.
- Hurd, Heidi. (1991). Challenging authority. *The Yale Law Journal*, 100(6), 1611-1677.
- Jacobson, Daniel. (2013). Regret, agency, and error. In David Shoemaker (Ed.), *Oxford Studies in Agency and Responsibility Volume 1* (95-125). Oxford University Press.
- Jaggar, Alison. M. (1983). *Feminist Politics and Human Nature*. The Harvester Press.
- Kain, Philip. J. (1993). Marx, housework, and alienation. *Hypatia*, 8(1), 121-144.
- Kant, Immanuel. (1996). Groundwork of the metaphysics of morals. In Mary Gregor (Trans., Ed.), *Practical Philosophy* (37-108). Cambridge University Press.

- Kant, Immanuel. (1996). The metaphysics of morals. In Mary Gregor (Trans., Ed.), *Practical Philosophy* (353-604). Cambridge University Press.
- Kiesewetter, Benjamin. (2022). Are all practical reasons based on value? In Russ Shafer-Landau (Ed.), *Oxford Studies in Metaethics, Volume 17* (27-53). Oxford University Press.
- Kimel, Dori. (2003). *From Promises to Contract*. Hart Publishing.
- Kimel, Dori. (2010). The morality of contract and moral culpability in breach. *King's Law Journal*, 21(2), 213-231.
- Korsgaard, Christine. M. (1983). Two distinctions in goodness. *The Philosophical Review*, 92(2), 169-195.
- Korsgaard, Christine. M. (1986). Kant's formula of humanity. *Kant-Studien*, 77(1-4), 183-202.
- Korsgaard, Christine. M. (2008a). The normativity of instrumental reason. In Christine M. Korsgaard, *The Constitution of Agency* (27-69). Oxford University Press.
- Korsgaard, Christine. M. (2008b). Acting for a reason. In Christine M. Korsgaard, *The Constitution of Agency* (207-229). Oxford University Press.
- List, Christian and Valentini, Laura. (2016). Freedom as independence. *Ethics*, 126(4), 1043-1074.
- Maguire, Barry. (2017). There are no reasons for affective attitudes. *Mind*, 127(507), 779-805.
- Marcus, Ruth. B. (1980). Moral dilemmas and consistency. *The Journal of Philosophy*, 77(3), 121-136.
- Markovits, Daniel. (2004). Contract and collaboration. *The Yale Law Journal*, 113(7), 1417-1518.
- Markovits, Daniel. (2011). Promise as an arm's-length relation. In Hanoch Sheinman (Ed.), *Promises and Agreements* (295-326). Oxford University Press.
- Markovits, Julia. (2010). Acting for the right reasons. *The Philosophical Review*, 119(2), 201-242.
- Marušić, Berislav. (2018). Do reasons expire? An essay on grief. *Philosophers' Imprint*, 18(25), 1-21.
- Marx, Karl. (1988). *Economic and Philosophic Manuscripts of 1844*. Trans. Martin Milligan. Prometheus Books.
- Mason, Elinor. (2019). Between strict liability and blameworthy quality of will: taking responsibility. In David Shoemaker (Ed.), *Oxford Studies in Agency and Responsibility, Volume 6* (241-264). Oxford University Press.
- McKenna, Michael. (2022). Reasons-responsiveness, Frankfurt examples, and the free will ability. In Dana K. Nelkin and Derk Pereboom (Eds.), *The Oxford Handbook of Moral Responsibility* (27-52). Oxford University Press.

- Mian, Emran. (2002). The curious case of exclusionary reasons. *Canadian Journal of Law and Jurisprudence*, 15(1), 99-124.
- Milam, Per-Erik. (2016). Reactive attitudes and personal relationships. *Canadian Journal of Philosophy*, 46(1), 102-122.
- Mill, John. S. (2015). *On Liberty, Utilitarianism, and Other Essays* (2nd ed.). Oxford University Press.
- Molina, Crescente. (2019). The authority in property. *Jersualem Review of Legal Studies*, 19(1), 14-20.
- Moller, Dan. (2007). Love and death. *The Journal of Philosophy*, 104(6), 301-316.
- Moller, Dan. (2019). *Governing Least*. Oxford University Press.
- Moore, George. E. (1922). *Principia Ethica* (1st ed., reprint). Cambridge University Press.
- Moore, Michael. S. (1989). Authority, law, and Razian reasons. *Canadian Journal of Law and Jurisprudence*, 15(1), 99-124.
- Moore, Michael. S. (1997). The independent moral significance of wrongdoing. In Michael Moore (Ed.), *Placing Blame* (1st ed.). Oxford University Press.
- Morse, Stephen. J. (2000). The moral metaphysics of causation and results. *California Law Review*, 88(3), 879-894.
- Movsesian, Mark. (2018). Markets and morals: the limits of doux commerce. *William & Mary Business Law Review*, 9(2), 449-475.
- Na'aman, Oded. (2021). The rationality of emotional change: toward a process view. *Noûs*, 55(2), 245-269.
- Nagel, Thomas. (1976). Moral luck. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 50(1), 115-135.
- Nelkin, Dana and Vargas, Manuel. (2024). Responsibility and reasons-responsiveness. In Taylor W. Cyr, Andrew Law and Neal A. Tognazzini (Eds.), *Freedom, Responsibility, and Value* (37-60). Routledge.
- Nussbaum, Martha. C. (1994). The *Oedipus Rex* and the ancient unconscious. In Peter L. Rudnytsky and Ellen H. Spitz, *Freud and Forbidden Knowledge* (42-71). New York University Press.
- Nussbaum, Martha. C. (2001). *The Fragility of Goodness* (2nd ed., reprint). Cambridge University Press.
- Oman, Nathan. B. (2016). *The Dignity of Commerce*. The University of Chicago Press.
- Owens, David. (2008). Rationalism about obligation. *European Journal of Philosophy*, 16(3), 403-431.
- Owens, David. (2012). The value of duty. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 86(1), 199-215.
- Owens, David. (2016). Promises and conflicting obligations. *Journal of Ethics and Social Philosophy*, 11(1), 1-19.

- Peels, Rik. (2015). A modal solution to the problem of moral luck. *American Philosophical Quarterly*, 52(1), 73-87.
- Penner, James. (2014). Promises, agreements, and contracts. In Gregory Klass, George Letsas and Prince Saprai (Eds.), *Philosophical Foundations of Contract Law* (116-134). Oxford University Press.
- Pepp, Jessica, Michaelson, Eliot and Sterken, Rachel. K. (2019). What's new about fake news? *Journal of Ethics and Social Philosophy*, 16(2), 67-94.
- Pereboom, Dirk. (2001). *Living Without Free Will*. Cambridge University Press.
- Perry, Stephen. R. (1992). The moral foundations of tort law. *Iowa Law Review*, 77(2), 449-514.
- Perry, Stephen. R. (2006). Second-order reasons, uncertainty and legal theory. *Southern California Law Review*, 62(3&4), 913-994.
- Pettit, Philip. (1997). *Republicanism: A Theory of Freedom and Government*. Oxford University Press.
- Piller, Christian. (2006). Kinds of practical reasons: attitude-related reasons and exclusionary reasons. In Sofia Miguens, João A. Pinto and Carlos E. Mauro (Eds.), *Analyses* (98-105). Porto University.
- Pinker, Steven. (2011). *The Better Angels of Our Nature: Why Violence Has Declined*. Penguin.
- Polanyi, Karl (1957a). Marketless trading in Hammurabi's time. In Karl Polanyi, Conrad M. Arensberg and Harry W. Pearson (Eds.), *Trade and Market in the Early Empires* (12-26). The Free Press.
- Polanyi, Karl. (1957b). The economy as instituted process. In Karl Polanyi, Conrad M. Arensberg and Harry W. Pearson (Eds.), *Trade and Market in the Early Empires* (243-269). The Free Press.
- Polanyi, Karl. (2001). *The Great Transformation* (2nd ed.). Beacon Press.
- Postgate, Nicholas. (2007). *The Land of Assur & The Yoke of Assur*. Oxbow Books.
- Rawls, John. (1999). *A Theory of Justice* (2nd ed.). Harvard University Press.
- Raz, Joseph. (1977). Promises and obligations. In Peter M. S. Hacker and Joseph Raz (Eds.), *Law, Morality, and Society* (210-228). Clarendon Press.
- Raz, Joseph. (1982). Promises in morality and law. *Harvard Law Review*, 95(4), 916-938.
- Raz, Joseph. (1986). *The Morality of Freedom*. Oxford University Press.
- Raz, Joseph. (1989). Facing up: a reply. *Southern California Law Review*, 62(3&4), 1153-1235.
- Raz, Joseph. (1994). *Ethics in the Public Domain* (1st ed., reprint). Oxford University Press.
- Raz, Joseph. (1996). *Ethics in the Public Domain*. Oxford University Press.
- Raz, Joseph. (1999). *Engaging Reason*. Oxford University Press.

- Raz, Joseph. (1999). *Practical Reason and Norms* (2nd ed., reprint). Oxford University Press.
- Raz, Joseph. (2001). *Value, Respect, and Attachment*. Cambridge University Press.
- Raz, Joseph. (2003). *The Practice of Value*. Oxford University Press.
- Raz, Joseph. (2004). Personal practical conflicts. In Peter Baumann and Monika Betzler (Eds.), *Practical Conflicts: New Philosophical Essays* (172-196). Cambridge University Press.
- Raz, Joseph. (2011). *From Normativity to Responsibility*. Oxford University Press.
- Raz, Joseph. (2021). Exclusionary reasons. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3933033.
- Regan, Donald. H. (1989). Authority and value: reflections on Raz's *Morality of Freedom*. *Southern California Law Review*, 62(3&4), 995-1095.
- Regan, Donald. H. (2004). Why am I my brother's keeper? In R. Jay Wallace, Philip Pettit, Samuel Scheffler and Michael Smith (Eds.), *Reason and Value* (202-230). Oxford University Press.
- Ripstein, Arthur. (2007). Tort law in a liberal state. *Journal of Tort Law*, 1(2), 1-41.
- Ripstein, Arthur. (2009). *Force and Freedom*. Harvard University Press.
- Rivera-López, Eduardo. (2016). How to reject resultant moral luck alone. *The Journal of Value Inquiry*, 50(2), 415-423.
- Rouse, Joseph. (2023). *Social Practices as Biological Niche Construction*. The University of Chicago Press.
- Salje, Léa. (2023). Depression, *ataraxia*, and the pig. *Australasian Journal of Philosophy*, 101(2), 251-266.
- Saprai, Prince. (2019). *Contract Law Without Foundations*. Oxford University Press.
- Saprai, Prince. (ms.). Trust in promise and contract: one sense or two? [Unpublished manuscript]. Faculty of Laws, University College London.
- Sartre, John-Paul. (2007). *Existentialism is a Humanism*. Trans. and eds. Carol Macomber and John Kulka. Yale University Press.
- Scalise Jr., Ronald. J. (2007). Why no "efficient breach" in civil law?: A comparative assessment of the doctrine of efficient breach of contract. *The American Journal of Comparative Law*, 55(4), 721-766.
- Schauer, Frederick. (1991). *Playing by the Rules*. Oxford University Press.
- Scheffler, Samuel. (2004). Projects, relationships, and reasons. In R. Jay Wallace, Philip Pettit, Samuel Scheffler and Michael Smith (Eds.), *Reason and Value* (247-269). Oxford University Press.
- Schurz, Gerhard. (2002). Ceteris paribus laws: classification and deconstruction. *Erkenntnis*, 57(3), 351-372.
- Seligman, Matthew. A. (2019) Moral diversity and efficient breach. *Michigan Law Review*, 117(5), 885-938.

- Sepinwall, Amy. (2017). Faultless guilt: toward a relationship-based account of criminal liability. *American Criminal Law Review*, 54(1), 521-570.
- Shelley, James. (2010). The default theory of aesthetic value. *British Journal of Aesthetics*, 59(1), 1-12.
- Sher, George. (2006). Out of control. *Ethics*, 116(2), 285-301.
- Shiffrin, Seana. V. (2012). Is contract a promise? In Andrei Marmor (Ed.), *The Routledge Companion to Philosophy of Law* (241-258). Routledge.
- Shiffrin, Seana. V. (2016). Enhancing moral relationships through strict liability. *University of Toronto Law Journal*, 66(3), 353-379.
- Shiffrin, Seana. V. (2017). The moral neglect of negligence. In David Sobel, Peter Vallentyne and Steven Wall (Eds.), *Oxford Studies in Political Philosophy, Volume 3* (197-228). Oxford University Press.
- Simmel, Georg. (1908). *Soziologie* (1st ed.). Duncker & Humblot.
- Simmel, Georg. (1950). The stranger. In Kurt H. Wolff (Trans., Ed.), *The Sociology of Georg Simmel* (402-408). The Free Press.
- Simmel, Georg. (1955). *Conflict and the Web of Group-Affiliations*. Trans. Kurt H. Wolff and Reinhard Bendix. The Free Press.
- Skinner, Quentin. (2010). *Liberty Before Liberalism* (1st ed., reprint). Cambridge University Press.
- Smith, Stephen. (2014). Strict duties and the rule of law. In Lisa M. Austin and Dennis Klimchuk (Eds.), *Private Law and the Rule of Law* (189-206). Oxford University Press.
- Snedegar, Justin. (2018). Reasons for and reasons against. *Philosophical Studies*, 175(3), 725-743.
- Sommers, Tamler. (2012). *Relative Justice*. Princeton University Press.
- Srinivasan, Amia. (2018). The aptness of anger. *The Journal of Political Philosophy*, 26(2), 123-144.
- Steele, Jenny. (2022). Rylands v Fletcher and strict liability. In Jenny Steele (Ed.), *Tort Law: Texts, Cases, and Materials* (5th edn) (715-749). Oxford University Press.
- Strawson, Peter. F. (2008). Freedom and resentment. In Peter F. Strawson (Ed.), *Freedom and Resentment and Other Essays* (1-28). Routledge.
- Stilz, Anna. (2013). Occupancy rights and the wrong of removal. *Philosophy & Public Affairs*, 41(4), 324-356.
- Sunstein, Cass. R. (2016) Beyond Cheneyism and Snowdenism. *The University of Chicago Law Review*, 83(1), 271-293.
- Sussman, David. (2018). Is agent-regret rational? *Ethics*, 128(4), 788-808.
- Svavarsdóttir, Sigrún. (2014). Having value and being worth valuing. *The Journal of Philosophy*, 111(2), 84-109.

- Theunissen, L. Nandi. (2018). Must we be just plain good? On regress arguments for the value of humanity. *Ethics*, 128(2), 346-372.
- Theunissen, L. Nandi. (2020). *The Value of Humanity*. Oxford University Press.
- Theunissen, L. Nandi. (2023). Explaining the value of human beings. In Sarah Buss and L. Nandi Theunissen (Eds.), *Rethinking the Value of Humanity* (225-247). Oxford University Press.
- Theunissen, L. Nandi. (2025). On the value of intellectual and aesthetic activity: a reply to Ross. In Robert Audi and David Phillips (Eds.), *The Moral Philosophy of W. D. Ross* (226-249). Oxford University Press.
- Thomas, Felicity, Wyatt Katrina and Hansford, Lorraine. (2020). The violence of narrative: embodying responsibility for poverty-related stress. *Sociology of Health & Illness*, 42(5), 1123-1138.
- Thompson, Michael. (2008). *Life and Action*. Harvard University Press.
- Tuomela, Raimo. (2012). Group reasons. *Philosophical Issues*, 22(1), 402-418.
- Varden, Helga. (2012). A Kantian critique of the care tradition: family law and systemic justice. *Kantian Review*, 17(2), 327-356.
- Velleman, J. David. (1999). A right to self-termination? *Ethics*, 109(3), 606-628.
- Velleman, J. David. (2008). Beyond price. *Ethics*, 118(2), 191-212.
- Vogt, Katja. M. (2017). *Desiring the Good*. Oxford University Press.
- von Wright, Georg. H. (1963). *The Varieties of Goodness* (1st ed.). Routledge & Kegan Paul Ltd.
- Walden, Kenneth. (2021). Review of *The Value of Humanity*, by L. Nandi Theunissen. *Ethics*, 131(4), 808-812.
- Wallace, R. Jay. (1994). *Responsibility and the Moral Sentiments*. Harvard University Press.
- Wallace, R. Jay. (2013). *The View From Here*. Oxford University Press.
- Watson, Gary. (1987). Responsibility and the limits of evil: variations on a Strawsonian theme. In Ferdinand Schoeman (Ed.), *Responsibility, Character, and the Emotions* (256-286). Cambridge University Press.
- Way, Jonathan. (2012). Transmission and the wrong kind of reason. *Ethics*, 122(3), 489-515.
- Weinrib, Ernest. J. (1995). *The Idea of Private Law*. Oxford University Press.
- Whiting, Daniel. (2017). Against second-order reasons. *Noûs*, 51(2), 398-420.
- van Wietmarschen, Han. (2021). Attitudinal social norms. *Analysis*, 81(1). 71-79.
- Williams, Bernard. (1976). Moral luck. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 50(1), 137-151.
- Williams, Patricia. (1991). *The Alchemy of Race and Rights*. Harvard University Press.

- Wittgenstein, Ludwig. (1969). *On Certainty* (1st ed.). Trans. Denis Paul and Gertrude E. M. Anscombe, Eds. Gertrude E. M. Anscombe and Georg H. von Wright. Blackwell.
- Wolf, Susan. (1980). Asymmetrical freedom. *The Journal of Philosophy*, 77(3), 151-166.
- Wolf, Susan. (1981). The importance of free will. *Mind*, 90(359), 386-405.
- Wolf, Susan. (2001). The moral of moral luck. *Philosophic Exchange*, 31(1), 2-16.
- Zimmerman, Michael. J. (2002). Taking luck seriously. *The Journal of Philosophy*, 99(11), 553-576.
- Zimmerman, Michael. J. (2004). Another plea for excuses. *American Philosophical Quarterly*, 41(3), 259-266.
- Ziporyn, Brook. (2009). *Zhuangzi: The Complete Writings*. Hackett Publishing.