

REMORSE: A PROSPECTIVE GENEALOGY

Elianna Gabriella Fetterolf

A thesis submitted for the degree of Doctor of Philosophy

Department of Philosophy, University of Sheffield

November 2014

The results, discussions and conclusions presented herein are identical to those in the printed version. This electronic version of the thesis has been edited solely to ensure conformance with copyright legislation and all excisions are noted in the text. The final, awarded and examined version is available for consultation via the University Library.

TABLE OF CONTENTS

ABSTRACT	4
ACKNOWLEDGEMENTS	5
INTRODUCTION	6
CHAPTER 1	13
1.1 Introduction	13
1.2 The morality system and its features	16
1.2.1 Moral obligation and its essentially practical character	18
1.2.2 The impossibility of moral conflicts.....	19
1.2.3 The “obligation-out, obligation-in” principle.....	20
1.2.4 The inescapability of moral obligation	21
1.2.5 Moral obligations are always overriding.....	22
1.2.6 Only an obligation can beat an obligation	22
1.2.7 Impersonality and purity.....	23
1.3 Constructing and charting MS-remorse	24
1.3.1 The impossibility of irresolvable moral conflicts.....	25
1.3.2 The disappearing ought and its comforting consequence	28
1.3.3 Tragic or lesser of two evils cases.....	31
1.3.4 Arriving at the distinction between MS-remorse and regret.....	33
1.4 Voluntary action, blame and remorse in the morality system	35
1.4.1 MS-remorse as the first-personal mirror of MS-blame	37
1.4.2 What counts as in an agent’s power?	38
1.5 What is the morality system’s purified conception of the voluntary?	40
1.5.1 The purified conception of the will in action	41
1.5.2 What does this picture of the simple and pure will do?.....	44
1.5.3 Where does the picture of the simple and pure will come from?	46
1.5.3.1 Williams’ genealogy of focussed blame	46
1.5.3.2 Williams on “the special fantasy of retrospective prevention”	48
1.5.3.3 Williams on retributive punishment and the moral law	49
1.5.4 Williams’ criticisms of this conception of the will in action	50
1.6 What, then, is MS-remorse?	52
CHAPTER 2	55
2.1 Introduction	55
2.2 Taylor’s project	57
2.2.1 Taylor on self-assessment.....	57
2.2.2 Taylor’s framework: a cognitivist view of the emotions.....	60
2.3 Taylor’s account of guilt	63
2.3.1 Acceptance of authority and Taylor’s conception of moral authority	64
2.3.1.1 Criticising the concomitant picture of guilt	65
2.3.2 Self-regarding guilt and the essential thought in guilt.....	67
2.3.3 Essential harm and non-essential harm.....	71
2.3.4 Causal responsibility	74
2.3.5 Repayment	79
2.4 Taylor’s account of remorse	81
2.4.1 Scope.....	82
2.4.2 Causal responsibility revisited.....	83
2.4.2.1 Taylor’s problematic Hamlet example	83
2.4.2.2 Remorse, guilt and causal responsibility	85
2.4.3 The view an agent takes of her action and its consequences.....	86
2.4.3.1 The artificiality of Taylor’s conception of regret.....	90
2.4.4 Reparation.....	91

2.5 Guilt and Remorse.....	93
2.6 Conclusion	96
CHAPTER 3.....	99
3.1 Introduction	99
3.2 Gaita on moral theory and some remarks on method.....	100
3.3 Gaita's account of remorse.....	104
3.3.1 The outward trajectory of remorse.....	104
3.3.2 Remorse as not mere feeling.....	107
3.3.3 The inward trajectory of remorse: radical singularity	109
3.3.3.1 The first-personal perspective	109
3.3.3.2 Common fellowship and radical singularity.....	111
3.3.4 How are the two trajectories in remorse related?.....	115
3.4 Corruptions of remorse.....	116
3.4.1 Are all forms of corruption in remorse forms of self-absorption?	119
3.5 Remorse and moral understanding.....	120
3.6 Blame, moral responsibility and remorse	124
3.6.1 Some mistakes of the narrow conception of moral responsibility.....	125
3.6.1.1 The case of Oedipus: broadening moral judgement from narrow blame.....	127
3.6.2 Gaita's broader conception of moral responsibility.....	129
3.7 Reparation	132
3.7.1 A misconception of remorse.....	132
3.7.2 Sketching reparation	133
3.7.3 The wrongdoer and forgiveness.....	137
3.8 Situating remorse in moral life.....	138
3.9 Conclusion	139
CHAPTER 4.....	140
4.1 Introduction	140
4.2 The standard interpretation of agent-regret.....	141
4.3 The errors of the standard interpretation.....	145
4.3.1 The forgotten voluntary cases of agent-regret.....	145
4.3.2 The scope of agent-regret.....	147
4.4 Undoing the morality system's dichotomy	153
4.5 What is the role of agent-regret in Williams?	156
4.6 Remorse by any other name?	161
4.7 Conclusion	165
CONCLUSION.....	167
REFERENCES.....	169

ABSTRACT

With the increased interest in moral psychology in the last fifty years, moral philosophy has seen an upsurge of work on moral emotions. More recently, much fruitful attention has focused on such emotions as shame, blame and forgiveness, often framed by a more general concern with whether and to what extent these emotions can be claimed to be distinctively *moral*. My thesis takes part in this increased attention to emotions but draws attention to what I take to be a significant neglect of a central, distinctively moral, emotion: remorse. Part of what explains the neglect, I argue, is a persistent underlying commitment to a narrow, purified conception of remorse. I take this idea of remorse to be most uniformly, and powerfully, expressed by what Bernard Williams calls “the morality system”.

I argue for a novel interpretation of the purpose of “agent-regret”, interpreting it as a transitional, ultimately disposable, concept—a stepping-stone concept that is needed to move away from the purified conception of remorse we inherit from the morality system. The overarching aim is to chart, critique and move away from the narrow conception in order to point to what a broader conception of remorse might look like. To make these moves in what I call a “prospective genealogy of remorse”, I centrally appeal to three contemporary accounts of remorse, one provided by Gabriele Taylor, another by Raimond Gaita, and (according to my exegetical argument) a third by Bernard Williams. Though we are currently at a mid-point in this forward looking process, the aim is to make the eventually broader conception of remorse more perspicuous but, equally, to show with increased clarity its importance for any moral philosophy interested in maintaining a plausible moral psychology.

ACKNOWLEDGEMENTS

For her exemplary commitment to truthfulness, her open-mindedness, her precision, her tirelessness and all-round extraordinary support, I would like to thank my supervisor, Miranda Fricker. Alongside being an incomparable role model, you have unwaveringly encouraged me to find and express my own voice, and, in managing to coax it out, you shall always have my immeasurable gratitude.

The thesis variously benefitted from the helpful questions and comments from graduate students and faculty members at Birkbeck during my time there. Also, I would like to thank Adrian Moore for his trenchant comments on the paper “Remorse Beyond the Morality System” that I gave at *The Moral and Political Legacy of Bernard Williams* conference at the University of Oxford (April 2014).

I would like to thank Mr. David Harvey for introducing me to philosophy and taking the time to encourage my passion for it. And to Stephen, I thank you for your anchoring insight, which helped me develop the confidence needed to complete this.

This thesis would not have been possible without the ceaseless anchoring love and support of my family. I want to thank Omi, and Ogi, for being the kind of grandparents you always want your parents to be but only realise in adulthood they could not be. To Linz, the most supportive sister I could ever imagine, for holding, and keeping steady, the belief well before I could, that this is something I really could accomplish. To Mark for always making me laugh, and for your constancy. And to Kat, for a friendship forged in early adolescent upheaval but has survived the last twenty-one years.

I want to thank Matthew for keeping me alive, and for always knowing the right time to dance.

And to mom, I want to thank you for teaching me to find the deepest pleasure in learning, to be fearless in exploration, and to be relentless in my curiosity.

INTRODUCTION

The last thirty years has seen the legitimization of moral psychology as an independently worthy domain of philosophical interest. Amongst the most noteworthy changes brought about by the increasing importance of moral psychology is the current philosophical attention to moral emotions. Blame, shame, guilt and forgiveness have all become significant sites of philosophical debate in moral philosophy. Despite these changes, I take there to be a striking, and important, omission in this developing landscape of moral emotion: remorse. One could understand the omission as demanding an argument for a vindicatory revival of remorse, and its place in the canon of valuable moral emotions. This, however, is not straightforwardly the character of what I am interested in pursuing. Rather, I am interested in the idea that, unlike the aforementioned moral emotions, conceptions of remorse (and the attendant phenomena) have received little to no critical analysis or questioning, whether in normative theories or in moral psychology.¹

Certainly, there must be reasons or explanations for why remorse has been philosophically neglected. It could, in part, be explained by its unequivocal status as a distinctively moral emotion. Namely, it is “given” that remorse is a moral emotion, such that, it does not require the kind of critical analysis other, less determinately moral, emotions demand. On the occasions remorse does appear in the literature, it is characteristically, and straightforwardly, understood as the first-personal reaction to voluntary breakings of moral obligation. In other words, it is typically understood as a (if not *the*) first-personal expression of moral responsibility. The emphasis, albeit with variations, tends to be on the idea that remorse is rationally, and appropriately, restricted to voluntary actions, and that its objects are those actions and consequences that typically attract justified blame. As I understand it, this picture of remorse as the first-personal expression of moral responsibility for

¹ I will be restricting my discussion to philosophical moral psychology. I will not be considering experiments or studies in empirical moral psychology. For a summary of empirical work in moral psychology see Doris and Stick (2005), and Doris (2010). For a recent book in philosophy that addresses the contemporary under-researched character of remorse in psychology, and to a lesser extent, in jurisprudence, see Proeve and Tudor (2010).

blameworthy actions (and consequences) that is rationally restricted to the voluntary is a veritable stronghold in contemporary moral philosophy.

However, even if one has no investment in changing this standard picture of remorse, few philosophers seem to even wonder what it is about remorse that makes it unquestionably *moral*. For my part, as this thesis will endeavour to illuminate, the answer lies in the distinctive connection remorse, as a first-personal expression of moral responsibility, bears to those harmed by one's actions and their consequences. But, if we accept the idea that remorse has a special connection with those harmed, and, that *this* is what anchors its moral status, then, as I will address throughout the thesis, this will affect the plausibility of the claim that remorse is restricted to the voluntary.

Further explanation for the lack of philosophical questioning of remorse is also offered by the pervasive pre-occupation with the third-personal in moral philosophy. Although the second-personal is now established, the first-personal has largely been marginalized, certainly in contemporary moral philosophy.² Where Stephen Darwall's (2006) work has been seminal in raising interest in the second-personal standpoint, I take Bernard Williams to be the central advocate for the importance of the first-personal in moral philosophy. This finds its clearest expression in his arguments against the representation of morality as fundamentally impersonal, and the excessively abstract character of contemporary moral theories. As Miranda Fricker eloquently explains, "much of Williams's philosophy can be read as working out the implications of the idea that ethical life flows from a radically first-personal source – what, as an individual, one cares about, what moves one, and what makes one's life worth living"(2009, 245).

In his capacity as advocate for the first-personal, Williams offers an important insight into how the first-personal is characteristically approached in moral philosophy. When the first-personal takes on philosophical relevance, consideration of it is often already conditioned by the demands of either a second-personal or third-personal standpoint. For example, as we shall see in Chapter 4, when Williams introduces "agent-regret" as a first-personal reaction

² This may be gaining some currency particularly with the recent introduction of R.Jay Wallace's *The View From Nowhere* (2013), which is focused on understanding, from a moral psychological standpoint, the retrospective dimension of the first-personal perspective.

to harm done, he alerts the reader to this tendency by making a methodological qualification: “The discussion is not in the first place directed to what we or others might say or think of these agents (though it has implications for that), but on what they can be expected coherently to think about themselves” (1981b, 27). Here, I take Williams to be asking the reader to *temporarily* suspend the ever-proximate question of how a victim and/or a third party may judge an agent and their actions in the types of cases he considers. Williams makes this qualification to focus our attention on what can be made of the kinds of thoughts and feelings a reasonable agent actually experiencing remorse or agent-regret might be thought to intelligibly have from their own, first-personal, perspective. The initial reason for this is to come to a description *from* the first-personal perspective, and to use this as his starting point.³

Williams’ interest in creating reflective space between the first-personal and the second- and third-personal in the territory of expressions of (moral) responsibility is unusual. However, it is an approach he shares with Gabriele Taylor and Raimond Gaita, whose accounts of remorse I also consider in this thesis. All three philosophers are interested first and foremost in coming to understand remorse from the agent’s perspective, as independently as possible from the question of how others might judge them. Moreover, each works from this starting point by taking the phenomenology of remorse seriously, and, using the thoughts and feelings of the experience of remorse as the basis for philosophical analysis. This style of approach presents a distinctive form of naturalism because it neither appeals to the ethology of other species in the style of neo-Aristotelian naturalists like Philippa Foot (2001) and Rosalind Hursthouse (1999), nor to experiments in psychology like John Doris (2005; 2010) and Stephen Stich (2005). Namely, it is a form of naturalism that is not reductive in the now standard ways. Of course, this invites its own set of tangled difficulties, particularly in relation to the familiar concerns of how a philosopher’s first-personal perspective can be appropriately objective or unbiased (leaving aside related issues of empirical accuracy without scientific

³ In his later work, Williams refers to this approach as involving “the ‘ethnographic stance’ [borrowing from Donald Davidson] and the possibility of understanding without identification” (Williams 2006d, 63). This perspective and its possibility of understanding without identification is central to what Williams calls “ideal anthropology” (2006d).

support). As we shall see, Gaita has a reply to some of these concerns, and I suspect that, in one direction, Williams would appeal to the impossibility of a view from nowhere in moral philosophy, and, in another direction, to the possibility of interpretation and understanding without identification (2006d, 62-64).

With these methodological continuities in mind, to begin the process of turning philosophical attention to remorse, I introduce the methodological framing device of a “prospective genealogy”. The purpose of a forward-looking, as opposed to conventional backward-looking, genealogy, is to diagnose where we are at present, to imagine where we might get to, and uncover what is available to us in order to move closer to that endpoint.⁴ Significantly, the ideal-bearing aspect of the imagined endpoint is not meant to suggest that there is some perfect, fixed point that we must arrive at. But, as we shall see in the example this thesis presents, nor is it that coming closer to our imagined endpoint is not something we ought to be interested, and invested, in doing.

Applying the approach of a prospective genealogy to remorse, I take the starting point to be the standard picture of remorse found throughout contemporary moral philosophy. Specifically, I identify this picture as, most basically, the conception of remorse found in what Bernard Williams calls “the morality system” (and a style of ethical theorising associated with it). Under the constraints of the morality system, particularly in connection with its purified conception of voluntary action, I argue that this picture of remorse is pressurized and distorted. To diagnose where we might be, and where we might get to with respect to moving away from the morality system’s narrow, purified conception of remorse, I consider the contemporary accounts of remorse offered by Gabriele Taylor, Raimond Gaita, and arguably a third by Bernard Williams. As I consider these accounts, the suggestion emerges that we are at a midpoint with respect to moving away from the morality system’s purified conception of remorse. Importantly, this midpoint is marked not specifically by the morality system’s purifying commitments but by what I take to be its increasingly doubtful insistence that remorse is rationally restricted to the

⁴ For Williams’ central discussion of the prospects of genealogy as a method for philosophical inquiry, see (2002, Ch.2).

voluntary. From here, I take it that we can begin to work out what an expanded form of remorse, liberated from the pressures and distorting constraints of the morality system, might come to. By the end of this thesis, it becomes clear how the prospective genealogy's structure helps to uncover important questions and puzzles relating to remorse and, at the same time, reveals routes of analysis and constructive materials we can take forward.⁵

The starting point for the prospective genealogy is the subject of the first chapter: the morality system's conception of remorse. Given that Williams did not explain what he took this conception to be, I first provide groundwork by presenting the critical features of the morality system. Once in place, I turn to constructing and charting its conception of remorse. To do so, I examine the morality system's picture of moral conflict, which reveals the restriction of remorse to broken obligations, and, its characteristic pairing with non-moral regret. But there is another significant restriction on remorse in the morality system: the voluntary. To explain what this comes to, I claim that the morality system implicitly upholds the idea that remorse, often taken as equivalent to "self-blame", is merely the first-personal mirror of second or third personal blame. With this claim in place, I explicate Williams' argument for how and why the morality system purifies the voluntary. The purified conception of the voluntary is then applied to the morality system's conception of remorse. The result, I argue, is a conception of remorse that is excessively self-involved, and ultimately attendant to an abstraction, "the moral law".

At the end of Chapter 1 a question is uncovered regarding how guilt, which is absolutely central to moral psychology, relates to remorse. As I understand it, this is an important question for anyone interested in understanding guilt or remorse. I use this question to frame the second chapter where I explain Gabriele Taylor's accounts of guilt and remorse—accounts that are variously representative of the morality system, and yet, in moments mark a stepping away from it. Taylor takes the view that we ought to understand guilt

⁵ This is not an *entirely* novel approach. David Russell (2013) argues that R. Jay Wallace's conception of moral responsibility is a version of the morality system's narrow conception of moral responsibility, and irrecoverably suffers from Williams' objections against it (Russell 2013, 196). From his argument Russell concludes that we ought to follow Williams in taking a broader ethical approach and, from there, uncover a broader conception of moral responsibility (for which we already have resources).

and remorse as properly distinct moral emotions that can be felt about the same event. Specifically, she takes guilt to be an emotion of self-assessment that is an essentially *self-regarding* response to moral wrongdoing. Remorse, she claims, is *not* an emotion of self-assessment, and that it is an essentially *other-regarding* response to moral wrongdoing. Broadly, I argue that Taylor goes wrong in failing to recognise that her account of remorse in fact shares its main features with her account of guilt. I diagnose her mistakes as arising from her picture of self-assessment, her various commitments to the morality system, and her systemic dualistic assumption that emotions can be either self- or other-regarding. Some of her difficulties, I suggest, would have been resolved if she recognised the possibility of an emotion being *relational*. Moreover, the problems of her account make available the natural thought that either there is simply no difference between guilt and remorse properly construed, or, that remorse is a form of guilt made distinctive by its special focus on victims (or harm done).

In the third chapter, I turn to an account of remorse outside the morality system in the novel Wittgensteinian account offered by Raimond Gaita. As I elaborate Gaita's account of remorse, various constructive alternatives to some of the weaknesses uncovered in Taylor's account, and, more generally, to the morality system's conception of remorse (and guilt), become available. A central line of constructive difference I pursue is that Gaita, unlike Taylor, does not split remorse and guilt. Instead, he argues that remorse, as a suffering recognition of one's guilt, has both self-regarding *and* other-regarding dimensions. On his view, genuine remorse is characterised by two somewhat opposing energies: it is at once essentially outward-looking in its concern with those harmed or wronged, and yet it is also depicted as radically inward-looking. As I explain, a view that affirms such opposing energies without synthesising them risks pressing remorse into an artificially dualistic mould—though this time the dualism is internal to remorse rather than generated by way of an alleged contrast with guilt, as it was with Taylor. To avoid the internal schism created by the opposing directions of attention that Gaita rightly sees in remorse, I argue that he needs to move beyond an account that merely affirms one then the other energy, and achieve instead a model that can synthesize them. He fails

to do this, and I argue that what is needed here is a model of remorse that depicts it as a *relational* emotion. Remorse, as I propose, does exhibit at once an outward and inward trajectory, and this is because remorse relates the subject and object of a wrongful action.

The fourth, and final, chapter focuses on undoing a ubiquitous though ultimately mistaken interpretation of Williams' term "agent-regret". Drawing on various accounts, I argue that there is a pervasive misinterpretation of agent-regret in the literature: that Williams coined "agent-regret" to refer exclusively, and paradigmatically, to those cases of purely accidental agency or bad moral luck cases. To show this to be mistaken, I analyse Williams' introduction of agent-regret in "Moral luck", which reveals that the standard interpretation cannot be supported at the level of the text. Aside from uncovering a widespread misinterpretation, the analysis provides the basis for my alternative interpretive claim that agent-regret is better understood as a *transitional* concept needed to move beyond the morality system in the territory of first-personal expressions of (moral) responsibility. If we understand "agent-regret" as a transitional "platform-like" concept, I argue that we can make sense of the idea of a prospective genealogy of remorse. But "agent-regret" is not just a theoretical device needed to move away from the morality system's distortion of the territory of first-personal expressions of responsibility. In the contents of "agent-regret", I suggest that we can also find constructive materials for a broader form of remorse. In this way, I take Williams to be providing a platform for philosophical reflection that can help us to consider what we might want from a form of remorse but, equally, what remorse beyond the morality system might actually come to.

CHAPTER 1

REMORSE IN THE MORALITY SYSTEM

1.1 Introduction

Despite a popular characterisation of Bernard Williams as a deeply pessimistic, wholly sceptical philosopher, I would like to present a rather different characterisation—one that recognises his work as interleaving destructive *and* constructive dimensions. Fuelling the predominantly destructive interpretation, much of the attention to Williams' work has been focused on his trenchant criticisms of utilitarianism, and Kantianism. And there is, of course, no question that Williams *is* deeply sceptical of these particular moral theoretical strongholds, as indeed he is sceptical of the whole practice of moral *theory* so prevalent in moral philosophy. Most moral philosophers, according to Williams, mistakenly take the business of moral philosophy to be the construction of moral theories, and are thereby susceptible to the simplification, reduction and false systematisation of what Williams takes to be the irreducible vagueness, diversity and conflict of human ethical life.

These tendencies, as manifest in the demands of moral theory, converge in interesting and complex ways with what Williams calls “the morality system” or sometimes “morality”.⁶ Williams takes the morality system to be a particularly narrow view of the ethical, with highly restricted conceptions of “moral obligation”, “blame” and “the voluntary”. It tends, according to Williams, to simplify and reduce these ethical concepts into strict, purified forms. In the process of narrowing and purifying ethical materials, the morality system participates in the kind of reduction and simplification found in moral theory. In this way, Williams takes there to be a close connection between the peculiar presuppositions and prejudices of the morality system, and the philosophical project of constructing moral theory so that different moral theories are seen as typically, and variously, manifestations or representations of the morality system (Williams 2006a, Ch.1; Russell 2013, 194-195). Moral theories, in other

⁶ I will use “the morality system” or “system”. For essays on the morality system see C. Taylor (1995, 132-155), Jenkins (2006, 53-86), and Louden (2007, 104-134).

words, often attempt to further simplify, reduce and systematise the already narrow morality system.

One of the key areas Williams takes the morality system to engage in its distorting pressurisation and purification is in the territory of crucial moral psychological emotions like blame, guilt and remorse (in the morality system, shame does *not* have the status of being “moral”). However, Williams does not adequately explain *how* the pressurisation and purification of them comes about and *what* exactly these emotions then come to in the system.⁷ This chapter will explain, in detail, how the morality system comes to purify remorse and, as we shall see, blame too. My principle aim in this chapter, then, is to uncover something that Williams does not, namely, the morality system’s purified conception of remorse (henceforth “MS-remorse”). This purified conception is effectively the starting point of an implied *prospective* genealogy of remorse, whereby we imagine the expanded form remorse might come to take if liberated from the pressures of the morality system.

In doing so, I am following Williams, Raimond Gaita, and Christopher Cordner in taking MS-remorse to be more or less representative of a common, though ultimately problematic, way of thinking about remorse (sometimes simply termed “guilt”) in moral philosophy. Its being “representative” and “a common way of thinking” is meant to support the possibility of it being more or less manifest across different moral theories. In this way, though it is rarely acknowledged as being constructive, Williams’ “morality system” creates an important conceptual platform for doing moral philosophical work, which can speak to many moral theories without commitment to any one of them.

At the centre of Williams’ attempt to move away from the morality system is his introduction of the broader notion of “the ethical”; hence, characterising the morality system as “a species of *ethical* thought”. Unlike the morality system which “demands a sharp boundary for itself” (2006a, 7), Williams lands on the following characterisation of *the ethical*: “However vague

⁷ This is not strictly a fair assessment as Williams does discuss what “guilt” can come to in relation to certain ideas found in the system in *Shame and Necessity* (1993a). However, even there, Williams does not sufficiently explain what guilt comes to in the system. At various points in this thesis, I will be using two aspects of that discussion: Williams’ brief characterisation of that picture of guilt, and his central criticism against it.

it may initially be, we have a conception of the ethical that understandably relates to us and our actions the demands, needs, claims, desires, and, generally, the lives of other people, and it is helpful to preserve this conception in what we are prepared to call an ethical consideration” (2006a, 12). More specifically, with regards to my purposes, he offers the following, “Any conception of the ethical will include in some form a concern for people directly affected by one’s actions, especially those to whom one owes special care” (Williams 1995e, 244). That this is presented by Williams as a feature of *any* conception of the ethical is very significant, in part, because it points to two related rarities in Williams’ work: (i) the isolation of something distinctly *ethical* (for, as part of his project of naturalisation in terms of “ideal anthropology”⁸, Williams is frequently occupied with showing how what is taken to be distinctively “moral” is not), and, (ii) it is as close to making a *universal* claim concerning the ethical as Williams gets.⁹

I will accept that Williams’ claim is correct. However, I will leave aside the partial reasons qualification, “especially those to whom one owes special care”, as this would quickly lead into familiar problems associated with explaining and justifying how this care can be extended to others beyond one’s “special sphere” (arguably questions at the centre of much moral philosophy, and its concerns with universal application). Instead, the focus will be on “a concern for people directly affected by one’s actions”, which can take different forms in different conceptions of the ethical. Following Williams, and exploring the morality system from the point of view of the ethical, I want to ask: what aspect of the morality system might count as capturing a concern for people directly affected by one’s actions? The answer one would expect, I take it, is that such a concern ought to find expression in morality in our first-personal reaction to wrongs we have done: that is, in remorse (or guilt). So, what characterises *that* conception of remorse, and does it capture a concern for people directly affected by one’s actions? This is a question I will also pursue into the next chapter.

⁸ See Williams’ (2006d) introduction of “ideal anthropology”.

⁹ There are other places, most obviously in Williams’ model of responsibility (1993a, Ch.3; 2006b; 2006d, 66-68).

To accomplish my aim, I first outline critical features of the morality system in order to elaborate Williams' stray remarks on MS-remorse. Once this platform is outlined, I consider the operations of MS-remorse by way of the impossibility of irresolvable conflicts in the morality system. In doing so, I uncover the system's key distinction between remorse and regret. With this in place, I go on to claim that the morality system casts remorse as the first-personal mirror of the characteristic second and third-personal reaction of the morality system, blame. Given that the morality system distils blame into a purified form that partners the similarly purified conception of the voluntary (as we shall see), this effectively purifies remorse to match. It, therefore, becomes relevant for the further explication of MS-remorse to consider the morality system's purified conceptions of the voluntary, and of blame (henceforth "MS-blame"). Taken together, these materials are the starting point of the prospective genealogy of remorse.

1.2 The morality system and its features

In Williams' arguably most famous contribution, *Ethics and Limits of Philosophy* (henceforth *ELP*), we find a *tour de force* critique of moral philosophy past and present. But, it is not merely that, for, we also find a masterful meta-construction of many of the features and presuppositions of a predominant style of ethical thought that Williams takes to pervade moral philosophy. In the final chapter of *ELP*, "Morality, A Peculiar Institution", Williams describes and critiques this collection of features and presuppositions as "the morality system". Prevalent in modern Western Christian culture, the morality system is a particular development and style of ethical thought that distinguishes itself in centralising, and conferring, special significance to its notion of "moral obligation" (Williams 2006a, 6). Significantly, and somewhat confusingly, "morality" should not be understood as merely an invention of philosophers, "It is the outlook, or, incoherently, part of the outlook, of almost

all of us” (Williams 2006a, 174).¹⁰ This style of ethical thought, and “its spirit, its underlying aims, and the general picture of ethical life it implies”, is so dominant that it occupies much of contemporary moral philosophy (Williams 2006a, 174). Vital to its dominance is the promotion of the mistaken view that this form of ethical thought alone constitutes or captures the whole of the subject of moral philosophy.

Although the morality system is a species of ethical thought, it is not a wholly determinate set of ethical thoughts. Rather, it “embraces a range of ethical outlooks” (Williams 2006a, 174). Kant, Williams explains, gives expression to the “purest, deepest, and most thorough representation of morality”, although most varieties of utilitarianism are also representatives of the morality system, but as marginal members (2006a, 174; 178). Such distinct moral theories, I take it, can be representatives of the morality system, in part, because they partake in some of the morality system’s purified or narrow conceptions of moral obligation, blame, remorse or guilt, voluntary action and moral responsibility. They also, typically, partake in some of the system’s presuppositions, and its demand for a sharp boundary between itself, taken as the whole of ethics, and everything else. According to Williams, this sharp boundary finds expression, for example, in a strong emphasis on a distinction between supposed “moral” and “non-moral” senses for words (2006a, 7; 174).

In light of these features, Williams thinks the morality system is “something we should treat with a special scepticism” (2006a, 6). For, though it can represent “a deeply moving picture of the ideal that human existence can be ultimately just”, the picture of ethical thought and life it presents is ultimately an illusion (Williams 2006a, 196). I agree with Williams that we should be sceptical of the morality system, particularly with respect to the false picture of ethical thought and experience it encourages. I will follow Williams in one of the central ways he argues against these ideas: to charge their moral theoretical representatives with upholding implausible and overly simplified moral

¹⁰ Brian Leiter (1997) argues that Williams is unlike Nietzsche insofar as Williams’ critique of morality is not directed to the living morality of his time but to living moral philosophers. But given that Williams clearly takes the morality system to be part of the outlook of almost all of us, Leiter is clearly mistaken.

psychologies, particularly with respect to the lack of significance conferred on the first-personal and, historically, on the partial (1993b; 2006a).

As I mentioned in introduction to the thesis, and as we shall see in Chapter 4, Williams can often be found looking to our lived emotional responses and experiences (moral and otherwise) to undermine the morality system's pictures of ethical thought, and moral concepts. To what extent should our philosophical understanding of our moral concepts like "remorse" relate to, and be reflective of, our actual moral responses and experiences? As I understand him, Williams takes it for granted that our moral concepts ought to be reflective of our experiences, however complex they may be. This question demands serious reflection for anyone interested in taking a non-reductive naturalistic approach to moral psychology. What it takes seriously will be variously explored across the thesis, and provide the basis for why I take the morality system's purified form of remorse, though powerful, to be ultimately problematic: it brings with it an inadequate conception of what the experience of causing harm to another can be. But, for now, I will explain the central features of the morality system: its special notion of "moral obligation"; the impossibility of moral conflicts; the *obligation-in*, *obligation-out* principle; the inescapability of moral obligation; the overriding-ness of moral obligations; the idea that only an obligation can beat an obligation; and, finally, the morality system's impersonality and purity.¹¹

1.2.1 Moral obligation and its essentially practical character

The morality system is essentially practical. What this means becomes clearer by focusing on its special notion of moral obligation. Moral obligation "is expressed in one especially important kind of deliberative conclusion—a conclusion that is directed toward what to do, governed by moral reasons, and concerned with a particular situation" (Williams 2006a, 174-175). In other words, moral obligation is expressed in a *practical conclusion*. If I find myself in a difficult situation where I am unsure what options I have for action or which

¹¹ Williams, of course, has objections, sometimes many, to all of these features across his work. Unfortunately, I cannot deal with them in any systematic way here, and will only be addressing a very few.

course of action to take, I deliberate about what to do, and decide that *all things considered* I should/ought to do *x* (where *x* is a course of action governed by moral reasons). When I conclude that all things considered I ought to do *x* I am, according to the morality system, expressing a moral obligation to do *x*. Moreover, when I conclude that all things considered I ought to do *x*, this is an expression of practical necessity: I *must* do *x*.

But, the morality system does not claim that all conclusions of moral deliberations are expressions of moral obligation. There can be conclusions of moral deliberation that take the form of *permissions* rather than obligations. Namely, the conclusion of one's deliberation can issue in the conclusion that one *may*, on this particular occasion, take the desired course of action. This is because, as it turns out, one is not under a moral obligation. In such cases, "I ask whether I am under an obligation and conclude that I am not" (Williams 2006a, 175). This, following Williams, points to two further ideas: (i) that, in deliberations, moral considerations can take the form of moral obligations, and (ii) the predominating presence, and priority, of moral obligation in deliberation. Namely, the idea of moral obligation, even when not expressed in a practical conclusion, is understood as the essential frame for moral deliberation as such.

1.2.2 *The impossibility of moral conflicts*

The essentially practical character of the morality system is evidently reflected in moral obligation being a kind of practical conclusion that will help to explain other related features. When an agent concludes that she ought to do *x* or is morally obliged to do *x*, she understands herself as under an obligation to perform a certain *action* (as opposed to, say, thinking or feeling something); "An obligation applies to someone with respect to an action—it is an obligation to do something" (Williams 2006a, 175). Yet, the action cannot just be any action, it must be something that is in the agent's power. This qualification, following Williams, is indicative of the morality system's adherence to the principle of *ought* implies *can*. For example, when I conclude that all things considered I ought to do *x*, *x* must be something it is in my power to do. If *x* is not in my

power in these circumstances, then I need to deliberate again (Williams 2006a, 175). Moral obligations, therefore, are not only actions an agent *ought* to do, they are also always actions an agent *can* do. Transposing the supposed special force of moral obligation into another version of the formulation produces: “what I am obliged to do *must* be in my power” [my emphasis] (Williams 2006a, 175).

As I understand it, the claim is advocating a generalised idea: that it cannot be the case that an agent is under a moral obligation in circumstances where she lacks the power to do the relevant action. When this feature is connected with what Williams calls “the agglomeration principle” (1973a, 180) another important feature of the morality system emerges: that moral obligations cannot really conflict (2006a, 176). The agglomeration principle says that, wherever I take myself to be obliged to do *x* and obliged to do *y*, then it follows that I am obliged to do *x*-and-*y*. But, I cannot do both *x* and *y*. If moral obligations are only actions an agent can do, then it follows from the fact that I cannot do both *x* and *y*, that at most one of *x* and *y* is a genuine obligation.

1.2.3 The “obligation-out, obligation-in” principle

So far, I have emphasised *particular* moral obligations as issuing from given moral deliberations on particular occasions. But the morality system also insists on *general* moral obligations. There is a pressure within the morality system towards generalisation that Williams calls “the *obligation-out, obligation-in* principle” (2006a, 181). As I understand it, the “obligation-out” refers to any particular moral obligation yielded from a given moral deliberation, whereas, the “obligation-in” represents a general obligation one is under independently of the particular circumstances. The “out-in” of the principle expresses the morality system’s view that any particular moral obligation can be explained as an instance of a general moral obligation. Any particular moral obligation, therefore, comes to require the logical backing of a general moral obligation, if it is to be a justified expression of obligation. If, for example, a good friend of mine is in a position where she needs help with something important, and I am one of a few people in a position to help her, I

may conclude that I am under an obligation to help her in these circumstances. That I take the immediate claim on me to be an obligation to help her is, according to the morality system, because I am, like others, under a quite general obligation to help those in need (Williams 2006a, 175; 181).

1.2.4 *The inescapability of moral obligation*

That everyone is typically under general moral obligations, directed at various moral objectives, can be seen to connect with yet another feature: moral obligation is inescapable (Williams 2006a, 177). I may acquire a moral obligation voluntarily, e.g., making a promise or entering a commitment, or, without having directly chosen it, e.g., as a doctor I am obliged to help others in certain ways and situations outside my professional practice. But, once I am under a moral obligation (particular or general) there is no escaping it: “moral obligation applies to people even if they do not want it to” (Williams 2006a, 178). Namely, when present, a moral obligation is binding irrespective of an agent’s desire to perform the required action and, even at the limit, irrespective of whether the agent “would prefer not to be in this system or bound by its rules” (Williams 2006a, 177).

The *sense* that moral obligation is inescapable, I take it, finds expression first-personally in the thought that, “what I am obliged to do is what I *must* do” regardless of whether it is an action I want to perform (Williams 2006a, 178). If I do not do what I morally *must* do, then I will be subject to the characteristic reaction of the morality system: blame. This is because “obligations have a moral stringency, which means that breaking them attracts blame” (Williams 2006a, 180). Equally, should I continue to see the action not taken as required of me, as a responsible agent in the system, I ought to feel remorse for having failed. The inescapability of moral obligation, I think, becomes clearer when one considers it in relation to those who want to live outside the morality system or not be bound by its rules. For, according to the system, not only can an outsider be under a moral obligation but, in failing to act in accordance with it, we can say that blame “would not be based on a misunderstanding” (Williams 2006a,

177). As Williams explains, “From the perspective of morality, there is nowhere outside the system, or at least nowhere for a responsible agent” (2006a, 178).

1.2.5 *Moral obligations are always overriding*

That a moral obligation can apply to an agent irrespective of their desire to perform the required action, and at the limit, irrespective of whether they want to be in the morality system at all, follows from the system’s view that moral obligations, when present, are always *overriding*, or *categorical*, considerations. This is the case regardless of whether the competing consideration is moral or non-moral. If I am deliberating between a moral consideration that takes the form of a moral obligation, and a consideration of any other form, the moral obligation will always trump the other consideration. If I am deliberating between a moral consideration that takes the form of a moral obligation, and a moral consideration that does not, the moral obligation will always trump the non-obligatory moral consideration. This is because, on the system, a genuine moral obligation, whenever present in deliberation, will already represent whichever course of action one has most moral reason to take. Another way to express this is to say that, whichever course of action one has most moral reason to take, necessarily takes the form of a moral obligation (Williams 2006a, 179). Therefore, according to the system, moral considerations that do not take the form of moral obligations or cannot be logically backed by a general obligation cannot really be competing considerations after all.¹²

1.2.6 *Only an obligation can beat an obligation*

If a moral obligation is the most stringent consideration in a deliberation about what to do, and defeats other considerations, the question is raised of what, if anything, can beat a moral obligation. The answer, according to the morality system, is a more stringent moral obligation: *only an obligation can beat an obligation* (Williams 2006a, 180). My friend, for example, is having a

¹² As I understand Williams, this results in an expansive force in the system, particularly in deontological versions of it, to accommodate the live diversity and range of considerations by turning as many considerations as possible into moral obligations (2006a, 179; 181-182).

difficult time, and I make a promise to help him move his belongings out of his ex-boyfriend's flat. The only thing that can cancel my promise without attracting blame is a more stringent moral obligation. So, if, on my way to meet my friend at the flat, I find myself in a situation where I am called upon to help someone in immediate danger, then I cannot be justly blamed for not fulfilling the promise to my friend. This is because something like a general moral obligation "to help others in an emergency" is seen to justifiably override the moral obligation to keep my promise.

1.2.7 Impersonality and purity

Finally, the last two features of the morality system I will briefly address are its impersonality, and its purity. Taking impersonality first, when engaging in deliberation about what one has most reason to do, the system demands of the agent that she abstract away from *who* she is. According to Williams, in Kantianism the agent is required to abstract in moral thought from her identity, and in utilitarianism she needs to disregard her separateness from others in order to focus her attention on producing the best consequences or state of affairs for the largest number of people (1981a, 3). The assumption here, I take it, is that the correct point of view for moral deliberation and action within the morality system is an impartial point of view, a view abstracted from an agent's particular life, projects and relationships. If the morally correct point of view requires an agent to abstract from her identity, her projects and commitments, this guarantees the impartiality of the morality system.

The impersonality in turn guarantees the purity of the morality system because it establishes conceptual space beyond the evident diversity and contingency of commitments any given individual or particular life has, which are bound up with considerations of luck. Considerations of luck can variously take the forms of outcome or resultant luck (i.e., luck in the way things turn out), constitutive luck (i.e., luck in who one is, in the traits and dispositions that one has), and circumstantial luck (i.e., luck in the surroundings one finds oneself in) (Nagel 1979). In requiring moral thought to abstract from *who* one is, it means that the morality system is protected against the vagaries of contingency;

it is a *luck-free zone*. Equally, in abstracting away from *who* one is, and, in particular, the constitutive luck that will have gone into making you who you are, the morality system attempts to correct for the potentially unjust, or at least not lucky, distribution of advantages or disadvantages. Williams writes,

The purity of morality itself represents a value. It expresses an ideal, presented by Kant, once again, in a form that is the most unqualified and also one of the most moving: the ideal that human existence can be ultimately just...The ideal of morality is as a value, moral value, that transcends luck. It must therefore lie beyond any empirical determination. It must lie not only in trying rather than succeeding, since success depends partly on luck, but in a kind of trying that lies beyond the level at which the capacity to try can itself be a matter of luck. The value must, further, be supreme. It will be no good if moral value is merely a consolation prize you get if you are not in worldly terms happy or talented or good-humoured or loved. It has to be what ultimately matters (2006a, 195).

The ideal of the morality system expresses a moral ideal: that human existence can be ultimately just. To guarantee this justice, the morality system maintains a picture of moral value that upholds it as transcending luck (beyond any empirical determination), and as a supreme value among all other values. As we shall see in section 1.5, this proves essential to the morality system's pressurization and purification of many of our moral psychological concepts.

1.3 Constructing and charting MS-remorse

Outlining these defining features of the morality system provides a platform from which I shall now try to construct an explicit understanding of what Williams takes MS-remorse to be. What emerges in connection with it is the morality system's distinction between MS-remorse and regret. Again, I will be using MS-remorse, and the associated distinction with regret, as a representative model of how remorse and regret are characteristically conceived of in moral philosophy (or, how the territory of first-personal expressions of responsibility is carved up). This, I take it, does not exclude the possibility that any particular version of remorse, and/or the distinction with regret, may not count as being more or less representative such that the criticisms are more or less relevant.

Before doing so, it is worth noting that, in *ELP*, Williams uses “remorse”, “guilt”, “self-blame” and “self-reproach” collectively or interchangeably when discussing the morality system’s first-personal expression of moral responsibility (2006a, 177). In the later *Shame and Necessity* (henceforth *SN*), Williams focuses his critique solely on guilt. To interpret his sometimes collective, sometimes interchangeable use of “remorse”, “guilt”, “self-blame” and “self-reproach”, I will draw on his comments regarding the relation between guilt and shame:

The mere fact that we have the two words does not, in itself, imply that there is any great psychological difference between shame and guilt. It might merely be that we set up an extra verbal marker within one and the same psychological field, in order to pick out some particular applications of what would otherwise be shame (Williams 1993a, 89).

Though Williams does, in fact, take shame to have a distinct psychological field from guilt, given that he does not distinguish between remorse, guilt, self-blame and self-reproach, I will interpret his interchangeable use in terms of what he says in the quotation. Namely, that there is no great psychological difference between them, such that we can say they are part of the same psychological field. Although they may belong to the same psychological field, I take it that self-reproach is broader than remorse because I may reproach myself for poor behaviour that actually helps someone, whereas I take remorse to be appropriate in cases where harm is done to others. In general, for my part, I agree with the idea that they are part of the same broad psychological field. How best to understand their relations within that field, however, is a question I shall keep exploring throughout this thesis.

1.3.1 The impossibility of irresolvable moral conflicts

I think a good place to begin in constructing and charting the operations of MS-remorse is to turn to the knotty issue of moral conflicts. Again, the morality system’s special notion of moral obligation is of a practical conclusion that is always overriding. When present in an agent’s deliberation about what to do, a moral obligation necessarily overrides other considerations, including

those which carry moral weight but are not or cannot take the form of a moral obligation. To clarify what exactly this comes to, Williams gives the example of promising to visit a sick friend in the hospital but getting the conflicting opportunity to significantly help an important cause (2006a, 180). “Helping out an important cause” does not readily take the form of a moral obligation, so following the system, it would seem that the promise, a prototypical obligation voluntarily undertaken, should necessarily outweigh the other moral consideration.

But, according to the system, should the moral consideration “helping out an important cause” be a genuine rival to the promise in deliberation, then this is only because it can take the form of an obligation. As we saw previously, in the system, the only thing that can beat a moral obligation is another, more stringent, moral obligation. Should it be the case that the moral consideration can take the form of an obligation, then there is a competition between two moral obligations, and it seems the agent lands in a conflict situation. The agent believes that she is required to keep her promise, and required to help the cause, but cannot do both. It seems that, no matter which course of action she chooses, she will fail to do something she is morally required to do. Consequently, in this situation, it appears inevitable that she will do something wrong or fail to do something she ought to.

However, as we saw above, in the system, moral obligations “cannot conflict, ultimately, really, or at the end of the line” (Williams 2006a, 176). Again, the feature of moral obligations that they cannot really conflict is taken to follow from the system’s adherence to the principle of *ought* implies *can*, and the agglomeration principle. The influence of *ought* implies *can* appears in the qualification that “what I am obliged to do must be in my power” (Williams 2006a, 175). If an agent concludes that she is morally obliged to perform a certain action, then she must logically or in principle be able to perform it (Williams 2006a, 174).

Any conflict example like the one above can be formalised in the following way: an agent in a particular situation believes that she is morally required to do *x*, and that she is morally required to do *y*. But, as this gives rise to the conflict, she cannot do both *x* and *y*, which will result in her not doing

something she ought to do or doing something she ought not to. When combined with the agglomeration principle the argument runs as follows (Williams 2006a, 176; 1973a, 179-184): I ought to do *x* and I ought to do *y*, therefore (by the agglomeration principle) I ought to do both *x* and *y* but (by *ought* implies *can*, used contrapositively) I cannot do both *x* and *y*, therefore it is not the case that I ought to do both *x* and *y*. This gives rise to a contradiction. To escape the problem, if we accept the sound conclusion that it is not the case that I ought to do both *x* and *y*, then it cannot be the case that I ought to do *x* and that I ought to do *y*. One of the “oughts”, therefore, must not apply.

It is supposed, according to Williams, that conflict is “a case of logical inconsistency, such that the agent could not be justified or rational in thinking that each of these moral requirements applied to him” (1981d, 74). However, Williams claims that this picture mistakenly gives rise to the idea that the source of trouble in a conflict situation lies with the agent’s thought about the moral situation as opposed to the trouble being located in the situation, which may or may not be the agent’s fault (1981d, 74).¹³ On the system’s picture, resolving the conflict is a matter of the agent being disabused of her mistaken view of the situation about which obligation does, in fact, apply (Williams 1973a, 175-176). In any given conflict situation where an agent believes that she ought to do *x* and that she ought to do *y*, but cannot do both without moral failure, it must be that she is mistaken, for example, in taking both of the claims to be genuine obligations.

On this view, conflicts are always merely apparent and eliminable. In a conflict situation, it must turn out that one of the obligations is more morally stringent than the other such that, one of them is seen to override or outweigh the other. Whether against some non-obligatory moral consideration or another moral obligation, what it means, in the system, for an obligation to outweigh or override is that when it does so, it eliminates its rival *completely*; “they [moral theorists] eliminate from the scene the ought that is not acted upon” (Williams 1973a, 184).

¹³ Following Williams, I am not considering those who are interested in arguing on metaphysical grounds that such a conflict situation is a metaphysical impossibility, or, in some way, a moral impossibility (1981d, 75; 1993a, 67-68).

The rejected obligation and the considerations or reasons which were, in deliberation, supporting it as a relevant course of action, rationally come to be seen as not applying; “it is a necessary consequence of conflict that one *ought* must be totally rejected in the sense that one becomes convinced that it did not actually apply” (Williams 1973a, 184). What I take from this is that the complete elimination of the conflicting obligation has the result that, wherever an agent acts in accordance with what she has most moral reason to do (i.e., the overriding moral obligation), *there can be no moral grounds for moral emotions like remorse or guilt*. The point is not merely that moral conflicts are always apparent or necessarily eliminable but also that they are necessarily resolvable without moral remainder or residue.

1.3.2 The disappearing ought and its comforting consequence

Regardless of how straightforwardly I come to the decision to act, in the morality system, it is the preference for one moral obligation over the other which rationally eliminates, or constitutes the total rejection of, one of the conflicting “oughts” from the scene (Williams 1973a, 184). “The evident fact that there is at most one of the two things which, all things considered, I should do, is taken to be equivalent to the idea that, all things considered, there is only one obligation” (Williams 1981d, 74). Once an agent decides that “all things considered she should do *x* and not *y*” it is, for example, as though it turns out *y* was not in fact an obligation, or, it may still be considered an obligation (i.e., a promise) but one that the agent was mistaken in regarding as ultimately applicable to the particular situation. Given that at the site of decision one of the obligations disappears, an agent who continues to uphold that the rejected action was required is irrational, or at least unjustified, in their thinking. This puts us in a position to identify an “MS-test” of whether an agent has successfully disabused herself of her “mistaken view” in a conflict situation: if she takes herself to have acquired a further obligation to compensate those affected by the neglected action, then she maintains that the rejected action did in fact apply.

As this MS-test suggests, this pattern of resolution where to end in decision is necessarily to eliminate one of the conflicting obligations has what Williams aptly calls a “comforting consequence” (2006a, 177). If the elimination of the conflicting obligation means that it did not apply, then there can be no course of moral action an agent failed to take. What this means it that there can be no moral grounds for justified blame or reproach from others, and, without any conception of moral remainder, no grounds beyond the justified reproach of others for remorse. “No actual obligation was broken. This has a comforting consequence, that I should not blame myself...it is mistaken to blame or reproach myself for not doing the rejected action: self-reproach belongs with broken obligations, and, it has turned out, there was no obligation” (Williams 2006a, 176-177). In light of my previous exegetical point regarding Williams’ interchangeable use of “remorse”, “guilt”, “self-blame” and “self-reproach”, I’ll take it to follow from this passage that the object of any one of these emotions is a broken obligation. Given that the apparent conflicting ought disappears, taking with it all of its moral considerations or reasons, MS-remorse can never be a rational response to moral remainders or residues (for, they don’t exist on the system). What this means for my construction, therefore, is that we have now homed in on the only proper object of MS-remorse: broken moral obligations.

If, as this account would have it, the conflicting item is eliminated and it turns out that there is only one obligation, then unless the agent fails to act in accordance with what all things considered she thinks she ought to do (i.e., she is subject to *akrasia*), or if she comes to think that she was mistaken in thinking that that was what she had most moral reason to do (i.e., a deliberative failure), there can be no object for rational remorse (or guilt). Echoing the MS-test, it follows that, should an agent continue to maintain that the rejected action was in fact required of her, then she may see herself as having broken an obligation, thus giving her moral grounds for rational remorse. This already suggests that, under the system, any first-personal reaction that may (morally and rationally) show concern for people directly affected by one’s actions is strictly delimited by what counts as the breaking of a moral obligation (Williams 1995e, 244). In other words, presaging what I will explore below, without a broken moral

obligation any rational concern for people directly affected by one's action must be classed as non-moral in kind.

At one extreme, it falls out of this picture of conflict that "if I am convinced that I acted for the best...then it is merely irrational to have *any* regrets" [my italics] which implies "that these reactions are a bad thing, which a fully admirable moral agent (taken, presumably, to be rational) would not display" (Williams 1973a, 173). If a fully admirable agent is rational, then were she in the position of the opening example, and decided that all things considered she ought to help with the cause, it must turn out that the promise to her friend did not apply. It may be thought, for example, that there was an implicit condition that, were something more important to come up, the agent would no longer be expected to keep her promise. Under this description, the agent cannot rationally be held responsible by herself (remorse) or others (blame) for not having acted in accordance with the promise, for there was no promise to break. The fully admirable agent should not feel any form of regret for not having visited her friend. Moreover, on this view, her friend will not have acquired the right to complain about the fully admirable agent's action, and the fully admirable agent will not be required to compensate her friend.

Although there will be cases where a conflict is quickly revealed as merely apparent, and cases where one obligation straightforwardly overrides another, following Williams, I take it that there may be others where decision might not come so easily. I take it that we can find ourselves in situations where one reasonably believes that one is under two stringent moral obligations but cannot do both, and thereby will fail to do something one ought to. In such a case, I may struggle to decide which course of action I think I have most moral reason to take. Since I actually need to take action, I may quite reasonably seek out (perhaps in the form of advice) a further moral reason or consideration in preference of one of the obligations. For example, "I may use some emergency provision, of a utilitarian kind for example, which deals with the conflict of choice, and gives me a way of 'acting for the best'" (Williams 1973a, 184). In light of this consideration, I decide that all things considered I ought to do *x* and not *y*.

The reality of such a situation, Williams argues, is such that, though I choose x , and not y , I may continue to hold that y was actually required of me, so that I acquire a further obligation or moral reason to compensate those affected by my not choosing y —to make amends in some suitable way to the person or people let down (2006a, 176). But, as we saw above, on the system, unless it turned out that I was mistaken about x , it would be irrational or unjustified to continue to hold that y did in fact apply in this situation, and demanded a moral response. From the point of view of the system, it is unintelligible to suggest that I am morally required to compensate those affected by my taking x , and not y . This, as we shall in the next section, comes under further critical pressure when we consider tragic or lesser of two evil cases.

1.3.3 Tragic or lesser of two evils cases

Tragic or lesser of two evil cases put the morality system's picture of conflict and its resolution under considerable strain. Following Williams, and other realists about conflict, I take it that, in such cases, "an agent can justifiably think that whatever he does will be wrong: that there are conflicting moral requirements, and that neither of them succeeds in overriding or outweighing the other" (1981d, 74).¹⁴ In these situations, it seems, remorse (or guilt) will be appropriate no matter what the agent does. Moreover, I take it that at least part of the reason we can take this to be the case is because the sorts of actions and consequences involved in these cases are amongst the most morally serious kinds of action. Some typical examples include: Sartre's case of the student who loses his brother in the war and must choose between staying at home with his mother as her only surviving child, or going to fight the resistance and risk losing his life; Agamemnon, who must decide whether or not to sacrifice his daughter to raise the winds enabling the fleet to set sail to fight at Troy; and Sophie, who, as an incoming prisoner at Auschwitz, is given an ultimatum: she has to choose which of her two children will be sent to a labour camp, whilst the other will be gassed.

¹⁴ The debate regarding the existence of moral dilemmas is, of course, long and complex. For some examples see Marcus (1980), Foot (1983), Greenspan (1983; 1995), Sinnott-Armstrong (1988), MacIntyre (1990) and for an interesting collection see Mason (1996).

To apply the system's logic of resolvable conflict to Sophie's situation will produce the now familiar results. To avoid having both children killed, Sophie deliberates and chooses to save her son's life by having him sent to the camp, thereby sending her daughter to her death. If we follow the system, when Sophie justifiably (e.g., because it is better that at least one child should survive) decides that all things considered she ought to save her son, rationally speaking, on the morality system, it must turn out that any moral requirement to save her daughter's life lapses. Again, as long as Sophie holds that all things considered she did what she ought to have done in saving her son, she should not feel any remorse with respect to the consequence of that choice: sending her daughter to her death.¹⁵

Generally, I take tragic or lesser of two evils examples to function well in connection with illuminating the absurdities of the system's results because they involve, as I said, the most serious kinds of moral actions and consequences. Also, they are often, though not always, partial in nature. In being so serious, and partial, I suspect they are better able to draw our attention to our lived moral sensibilities, which may not be so easy to contact in the more everyday conflict cases.¹⁶ For my present purposes, these cases can help bring into clearer focus what the morality system's commitment to the non-existence of moral costs or remainders in connection with moral response amounts to:

Morality resists the notion of a moral cost, in the sense of a moral wrong knowingly committed by an agent who is doing something that even from a moral point of view is better: in that case, it will say, the wrong cannot ultimately be a wrong, the cost cannot really be a moral cost (Williams 1995e, 246).

This is clear in Sophie's case. To suggest, as the system's picture of conflict does, that any moral requirement she has to protect her daughter's life does not apply in light of some more stringent moral requirement, i.e., saving any life, produces

¹⁵ Williams makes the point that, in such cases, it may well be that "the notion of 'acting for the best' may very well lose its content", and that, "The agonies that a man will experience after acting in full consciousness of such a situation are not to be traced to a persistent doubt that he may not have chosen the better thing; but, for instance, to a clear conviction that he has not done the better thing because there was no better thing to be done" (1973a, 173).

¹⁶ In this connection, Williams writes, "it [moral costs] is a notion deeply entrenched in many people's moral consciousness. Why so many moral philosophers learn to forget it is a harder question, and perhaps a deeper one, than why some politicians do" (1981c, 63).

the result that, under those circumstances, knowingly sending her daughter to her death is not a morally relevant action. What would, in other circumstances, be considered a stringent moral obligation becomes, under the logic of guaranteed conflict resolution in the morality system, irrelevant, except insofar as it facilitates correct moral action. Sending her daughter to her death is an agony, though not ultimately a *moral* cost in the course of acting for the best. On this view, she may feel naturally distressed at sending her daughter to her death whereas any moral response, like remorse (or guilt), would be strictly irrational—an understandable confusion of natural response and moral response.

1.3.4 Arriving at the distinction between MS-remorse and regret

An objector to this view, like myself, may reasonably charge it with being disconnected from our ethical experience, which anticipates that Sophie would feel the deepest remorse at having to sacrifice her daughter. In fact, many are likely to take it as a signal of her moral decency that, in such a situation, she would feel the deepest remorse. Less rigid accounts continue to avoid this point, however, they do tend to acknowledge the psychological implausibility of the more draconian accounts. It is conceded that an agent—even on occasion the fully admirable moral agent—may understandably feel badly in such tragic situations (Williams 1973a, 173-174).¹⁷ It is conceded that the agent, like Sophie, may reasonably feel badly at having to do something disagreeable or distressing in the course of acting. “[I]f he has caused pain, in the course of acting (as he sincerely supposes) for the best, it might be said that any regret or distress he feels about having caused the pain is independent of his views of whether in doing this, he did something that he ought not to have done: he is just naturally distressed by the thought of having caused pain” (Williams 1973a, 175). This, I take it, is a concession to the reality of these situations—bad feelings may reasonably exist, but they cannot be moral in kind.

Without the presence of a broken obligation, so the system goes, these bad feelings related to one’s action cannot be classed as moral feelings. Rather,

¹⁷ See Baron (1988) for an essay attempting to reconcile these dynamics in the case of the ideal virtuous agent.

they are to be classified as natural responses to disagreeable or distressing deeds in the course of morally right action. So, on this concession, Sophie can understandably feel badly, even deeply badly, about having to send her daughter to her death, to do what she ought to, but these bad feelings are nevertheless distinguished from the moral feelings of remorse (or guilt) as “regret”, a non-moral feeling (Williams 2006a, 176-177).

The system responds to the pressure of the reality of bad feelings in such serious situations, without compromising its picture of the rationality of moral deliberation, by creating this distinction between remorse and regret as moral and non-moral responses to one’s actions. The system takes their differences to be reflected in their associated thoughts. If, for example, Agamemnon sincerely supposes that, all things considered, he did what he ought to have done in enabling the fleet to set sail, then this is incompatible with what we are now in a position to clearly identify as the thought in MS-remorse—e.g., “all things considered what I did was wrong” or “all things considered I did what I ought not have done”. Again, an agent can come to think this retrospectively in light of something like *akrasia* or deliberative error.

If Agamemnon, for example, decided that all things considered he ought to help the fleet, and yet, at the moment of sacrifice, found that he could not kill Iphigenia, so long as he continued to suppose that acting in accordance with his obligation to the fleet was what he ought to have done, he would have grounds for thinking that all things considered he acted wrongly and, thereby, feel remorse. Another variation has similar results: should he change his mind after killing Iphigenia and come to see preserving his daughter’s life as what he ought to have done, he would then see his initial view as mistaken, and his actions as constituting moral wrongdoing paving the way for remorse. Where in the former scenario, his remorse can be connected with a failure at the moment of action to do what he ought to have done, in the latter, we might say that he made a deliberative error, and if he had the opportunity to make the decision over again, he would make it differently (Williams 1973a, 185).

But, if Agamemnon does act in accordance with what he (justifiably) supposes he all things considered ought to do, and continues to believe this after killing his daughter, any bad feelings he has with respect to the killing are

to be understood as non-moral regret. His bad feelings or natural distress would then be characterised not by the MS-remorse thought “all things considered what I did was wrong” but by something more like the thought of regret, “how much better if it had been otherwise”. This thought, following Williams, is “available to *anyone* who can form some idea of how it might have been otherwise plus consciousness of how things would then have been better” (1981b, 27).

The system’s introduction of the distinction to preserve its picture of rational deliberation, at the same time as acknowledging the reality of potentially disagreeable aspects of morally right action, is reflected in the compatibility of the characteristic thought of morally right action, and that of regret. The characteristic thought of morally right action, “all things considered I did what I ought to have done”, is clearly compatible with the generic thought of regret, “how much better if it had been otherwise”. Agamemnon, for example, can compatibly think that he did what he ought to have done, but also, that it would have been much better if he had not had to kill Iphigenia in order to do so. As I understand it, this non-moral feeling of regret is the same regret, though sometimes qualified as “deep”, across all cases no matter how harmful or distressing the actions involved in doing what one ought to do. That is, its strength does not change its classification as non-moral. Moreover, the fact that this non-moral regret can be felt by anyone who knows of it, that it does not require the perspective of causal responsibility (“*I did it*”) suggests that this feeling of regret is further distinguished from remorse for, as I will take up in the next section, remorse is always an expression of agential responsibility.

1.4 Voluntary action, blame and remorse in the morality system

The focus I have so far taken on the mechanics of moral conflict in the morality system has helped to outline some key features of MS-remorse: its object is always a broken moral obligation, and whether or not the agent was in a conflict or non-conflict situation, in order for her to experience remorse, she must sincerely believe that she failed to do something she ought to have done or done something she ought not to have done. She must believe that “all things

considered what I did was *wrong*". Moreover, I uncovered how MS-remorse comes to be part of a central distinction with non-moral regret.

Now, I can say that the scope of MS-remorse is delimited by broken moral obligations. But, as I will explicitly explore here, it is also delimited by the condition of *voluntary* action. The question of the rationality of remorse depends not simply on whether or not I have done what I ought to, but whether the action was voluntary or involuntary. "The point that self-blame or remorse requires one's action to have been voluntary is only a special application of a general rule, that blame of anyone is directed to the voluntary" (Williams 2006a, 178). The rule that blame of anyone is directed to the voluntary (i.e., "the doctrine of the voluntary") is, according to Williams, an expression of the morality system's aspiration that blame "should apply only to the extent that the undesired outcome is the product of voluntary action on the particular occasion" (2006a, 194).

This rule, and its aspiration to focus rational blame exclusively on voluntary action, is connected with that fundamental commitment of the system to justice in moral responsibility (Williams 2006a, 177-178; 193). If, as the morality system would have it, moral obligation is both inescapable *and* in an agent's power to perform or refrain from performing (*ought* implies *can*), then transgressing a moral obligation will attract blame. The strictness of the inescapability of moral obligation is, to some extent, mitigated by the system's commitment to justice in responsibility: it is taken as a "requirement of justice that the agent should be blamed for no more and no less than what was in his power" (Williams 1995d, 72). The thought that an agent should be blamed or held morally responsible for no more and no less than what was in his power outlines a guarantee against being held morally responsible for instances of unavoidable wrongdoing or bad moral luck. Namely, it can provide security against unavoidable or tragic moral failure—recall, the comforting consequence in conflict situations. The morality system secures agents (first-personally and third-personally) against bad moral luck by making it unjust to hold an agent morally responsible for actions that were *involuntary* (or not in an agent's power).

1.4.1 MS-remorse as the first-personal mirror of MS-blame

What the restriction of the voluntary secures is the basis for my introduction of an idea that motivates the remaining sections of this chapter. This is an idea, which Williams never explicitly expresses, but I take to be implicit in his account: MS-remorse is conceived of as straightforwardly the first-personal mirror of MS-blame. Namely, in the system, it is assumed that blame is mirrored in the self, to get “self-blame”, which means that remorse, on this view, is the same as “self-blame”. This mirroring is suggested by blame’s status in the system as *the* characteristic second- and third-personal reaction to the breaking of moral obligations, and remorse’s status as being *the* characteristic first-personal reaction to the breaking of moral obligations. It is also supported by the equivalent relevance of the rule of the voluntary to both reactions. It does not seem, therefore, too far a stretch to presume that the application of the rule requiring that MS-remorse be restricted to voluntary action will be similar to, if not identical to, the case of MS-blame. Applied directly, and from the perspective of the morality system, I take it the result would be: it would be unfair to hold oneself morally responsible for the breaking of a particular obligation if the action was involuntary.

Transposed to the level of the individual agent, this is the idea that it would be irrational to feel remorse for breaking a particular moral obligation if the action was involuntary. Varying Williams’ famous lorry driver case, a train driver pulling into the station hits and kills a passenger who suddenly, and accidentally, falls onto the tracks. The moral obligation here is a general one, something like, “we should not kill others”. There is no way the driver could have put the brakes on in time, nor was he in any way negligent (i.e., falling asleep at the wheel or on mind altering drugs), and nor was the event foreseeable. The idea is that to blame the train driver for killing the passenger, and hold him morally responsible for the accident and its tragic outcome, would be unjust. Equally, it would be irrational, on this view, for the train driver to hold himself morally responsible, for, the only psychological formulation available on the morality system would be its conception of remorse, but one cannot rationally feel remorse in such a case of unavoidable wrongdoing (or bad

outcome moral luck). As we saw previously, he could respond to the death of the passenger with natural distress, and non-moral regret but, as will be discussed in Chapter 4, this is not an expression of *responsibility*, for anyone (the train driver, another passenger) who knows of the accident can regret it.

1.4.2 What counts as in an agent's power?

In its commitment to the just application of MS-blame and MS-remorse to voluntary action, the morality system, of course, faces the question of what counts as in an agent's power. It is "notoriously problematical", according to Williams, "not only because of large and unnerving theories claiming that everything (or everything psychological) is determined, but also because it is simply unclear what it means to say that someone can act, or could have acted, in a certain way" (2006a, 175). Here, there are two reasons that relate, albeit differently, to the problem of determining or deciding when something is or was in the agent's power, and when it is or was not in the agent's power. The first concerns the preoccupations of the free will debate, and the "large and unnerving theories" of determinism whether causal/physical, psychological, biological or theological (Williams 2006a, 175). These theories of determinism variously uphold the view that everything that happens (in the world) is causally necessitated by what comes before it. If this is correct, then it seems to conflict with the (typically metaphysical) requirements of the freedom of the will and moral responsibility.

On this view, for agents to be genuinely morally responsible for their actions, their will, the faculty by which an agent displays power or control in deciding and initiating action, must be free. If causal determinism is true, then it seems that every act of human will is itself necessitated by what comes before it, which means that the will is not free in the required sense for genuine moral responsibility. This can give rise to a picture of action in which the human will seems more "like a weathervane on a well-oiled pivot in a changeable wind which points in whichever direction it is blown" (Fricker 2009, 257). Namely, that the will, the faculty by which a person decides on, and initiates action, is not properly regulated by the agent herself. Without this possibility, so the worry

goes, there can be no proper object for moral responsibility (i.e., no agent in action).

The second reason, that it is unclear what it means to say that someone can act, or could have acted, in a certain way, relates to the interpretive openness, and indeterminacy of what it is for something to count, or have counted, as in an agent's power. It might be unclear, for example, what it means to say that someone can act or could have acted in a certain way because of considerations of luck, particularly in terms of constitutive luck. If a young woman has grown up with parents who are drug addicts, there is an important sense in which, should she become an addict herself, and support her habit through theft, one might wonder whether she really can act otherwise or, if she can, what it might mean to say that she can or could have in the past. We might maintain the thought that she can or could have acted otherwise but, in another sense, that she could not have. In a related direction, if I am being tortured for information that I do in fact have, it can seem to be both in my power to stop the torture, and not in my power to stop the torture. One might say that "it's not *really* in my power" because I am not being presented with a *real* choice.¹⁸

This question of what is and what is not in an agent's power is, of course, a general problem. But, for the morality system, it takes on an acute form. According to Williams: "the morality system lays particularly heavy weight on the unsure structure of voluntariness" (2006a, 243). As we saw above, in order to uphold a certain conception of justice in responsibility, the morality system lays heavy weight on the notion of voluntary action. "There is a pressure within it [the morality system] to require a voluntariness that will be total and will cut through character and psychological or social determination, and allocate blame and responsibility on the ultimately fair basis of the agent's own contribution, no more and no less" (2006a, 194). Here we have the two interrelated major objections Williams makes against voluntariness in the system: (i) the cardinal role the system assigns to voluntariness in its structuring of our moral psychological responses, and relatedly, (ii) the conception of voluntariness the system advocates. Though interrelated, I will be addressing them separately, the

¹⁸ Williams takes such a situation to be a red herring with respect to understanding free will and choice. Such cases, he argues, are better understood as examples of choice but of choice made under particularly narrow constraints (1995a, 5).

latter in the upcoming section, whereas the former will take on greater significance in Chapter 4.

As I read Williams, character and psychology or social determination are entangled with difficult questions of constitutive luck and, morality being immune from external contingency (i.e., its purity), needs to find a species of action that is free from luck. It also needs an understanding of an agent's power that is free from external contingency or empirical determination, such that blame and moral responsibility are allocated fairly on grounds that are non-contingent. The morality system has a pressure, or expectation, built into it that there can be such a thing as the determination of an agent's own contribution to an action that is free from considerations of empirical determination.

This pressure, understood as a certain freedom from luck guaranteed by an isolation from empirical determination, supports an ideal of a *purely* voluntary act that is closely connected with, and necessary for, moral responsibility (in the morality system) to obtain. This ideal is supported by a picture of action the system emphasises in its concentration on a "purified conception of the will" (Williams 1995a, 16). As we shall see in the next section, according to Williams, it is this picture of action in terms of the will that purifies MS-remorse—the system's first-personal expression of moral responsibility for the voluntary breaking of a moral obligation, i.e., wrongdoing.

1.5 What is the morality system's purified conception of the voluntary?

Across his work, Williams is frequently critical of the morality system's purified conception of voluntary action. In line with his criticism of the idea of moral value as a species of value that transcends luck, Williams claims that the purified conception of voluntary action, along with its cognate conceptions of purified blame, guilt, and moral responsibility, is an illusion (2006a, 196). Part of what makes these purified conceptions illusory, according to Williams, is their dependence on a peculiar (moral) psychology—a purified conception of the will in action. Briefly, this yields the following description of the system's conception of voluntary action: as a "morally significant interpretation of action in terms of the will" (Williams 1995d, 70).

Despite the prevalence of his dissent against it, there is only one place where Williams elaborates what he takes the purely voluntary to be—in his paper, “Nietzsche’s Minimal Moral Psychology” (1995d). I will follow Williams in first outlining a particular yet authority bearing (in philosophy and our actual moral psychology) interpretation of action in terms of the will that is taken as paradigmatically voluntary action. To do so, I elaborate the purified conception of the will in action. I then consider, following Williams, what this picture of action does, and where it comes from. Briefly, in answer, Williams claims that by observing the direct fit between this picture of action in terms of the will, MS-blame, and some of the system’s demands, it becomes evident that this picture of voluntary action (this special piece of psychology) “is itself a moral conception, and one that shares notably doubtful features of that particular morality itself” (1995d, 74). As we will see in section 1.5.4, in his critique, Williams takes this picture of voluntary action to be neither universal, nor coherent.

1.5.1 The purified conception of the will in action

To get to the point where it is clear that this picture of action in terms of the will is in service of the morality system (i.e., it is *its* conception of purified voluntary action), I will start where Williams does in building up a critical account of this authority bearing interpretation of action. Williams follows and elaborates some guiding diagnostic points from Nietzsche regarding which ideas typically go into this interpretation of action in terms of the will. “The belief in the will involves, for him [Nietzsche], two ideas in particular: that the will seems something simple...and that what seems simple also seems to be a peculiar, imperative, kind of cause” (Williams 1995d, 70). Taking the first idea, according to Williams’ Nietzsche¹⁹, to suppose that the will is something simple is a mistake. It is a mistake that arises, in part, because of an illusion perpetuated by “grammatical habit” —the will is “something that is a unit only as a word” (Nietzsche 1966, 19). Turning one’s attention to the experiences involved in willing, “a complex of sensations, thinking, and an affect of

¹⁹ In what follows, I will use “Nietzsche” as shorthand for “Williams’ interpretation of Nietzsche”.

command” (Williams 1995d, 70), suggests that the will is not something simple, a unit, but is complicated.

The belief in the will Nietzsche is attacking, this supposedly simple thing, also involves the idea that the will appears to be a “peculiar, imperative, kind of cause”. How does this idea of the will as a peculiar, imperative kind of cause arise? Turning to Nietzsche’s description of the experience of willing, of commanding oneself to obey one’s command in the successful issuing of an action, and how this perpetuates the picture of action under attack, Williams quotes the following passage:

Since in the great majority of cases there has been an exercise of will only when the effect of the command – that is, obedience; that is, the action – was to be expected, the appearance has translated itself into the feeling, as if there were a necessity of effect. In short, he who wills believes with a fair amount of certainty that will and action are somehow one; he ascribes the success, the carrying out of the willing, to the will itself, and thereby enjoys an increase of the sensation of power that accompanies all success (Nietzsche 1966, 19).

The point is not, Williams asserts, that the experience of willing “sets itself up as a sufficient cause of an action” (1995d, 70). The point of the passage, rather, is to reveal an illusion that emerges from the successful experience of willing: that there is a unique kind of cause, an “agent-cause”, that has a necessary relation with its effect, an action.

As I understand it, when an agent decides that she would like to do something, or she commands herself to do something, and it results in the required/desired action, she forgets the duality of her being both the commanding and obeying party. Namely, she forgets that she carries out the command and the willing, and believes her action is the straightforward product of her “simple” will. In light of the feeling of power that follows from the successful carrying out of the willing, she retrospectively takes her will to be a cause that necessarily produces its desired effect, an action. In this way, action is understood as something that requires a self or person or agent with a will to cause it.

Following Williams, there is a further dimension to this picture of the will as a unique, imperative kind of cause: that it “does not lie in any event or

state of affairs—whether an experience of mine or otherwise—but in something that I refer to as ‘I’ (1995d, 70). Splitting this quote, we have two important thoughts. The first is the idea that the imperative kind of cause, the “simple” will, lies in something that I refer to as “I”. The second idea, related to the first, is that this peculiar kind of cause exists independently of any causal (whether internal or external) set of events or states of affairs. Together these ideas suggest a further notion: that the thing, the “I”, in which the simple will lies should itself be understood as independent of any causal set of events or states of affairs, i.e., as free from empirical determination.

If the simple will is not fundamentally influenced in action by external causal events it may escape the worries of causal determinism. And, if the will is free from internal states of affairs, the picture helpfully avoids the difficulties associated with constitutive luck mentioned previously. Moreover, given that this peculiar cause only relates to its effect by way of prescription, through a command, which, when combined with its freedom from the influences of states of affairs or causal events, this suggests that the will, the site of the sensation of power, can produce its effect, an action, out of nothing. So the connection between the will and action seems to be a unique kind of necessary relation: it appears to be absolute. Namely, an action is not a mere reaction to external stimulus by an agent, it is necessarily the unmediated product of the will.

With this picture, according to Williams, the association between the following two ideas is evident: “that there is a metaphysically special unit, a real action, unlike anything else that can be individuated among the world’s processes”, and that *real action* stands “in an unmediated relationship – something like being an effect *ex nihilo* – to something of quite a different kind, again unique – a person or self, or agent” (1995d, 72). This conceptualisation of the will, and its special relation to “real” action, avoid the contextual difficulties and messiness of determining or deciding what counts as in an agent’s power. It does so by guarding the will against exposure to the contingencies of luck and empirical determination. There is, on this picture, a site of power in agency that is free from empirical determination and isolated from the vagaries of luck: the will. This is what is meant by describing the morality system’s picture of the will as simple but also as *pure*.

1.5.2 What does this picture of the simple and pure will do?

This special picture of action is, according to Williams, not just a philosophical theory but also a living piece of our moral psychology that can be reflected in our thoughts and moral reactions (1995d, 71). Given this, Williams takes this special picture of action in terms of willing as warranting further inquiry. He asks where it comes from, and what it does (Williams 1995d, 71). Taking the latter question first, Williams follows Nietzsche's explanation in the provision of two guiding thoughts on what this picture of action does, and what its purpose might be. The first helpful thought is that this picture of action "involves a kind of double counting" that happens at the level of entities and actions/what is happening (Williams 1995d, 71). "The self or I that is the cause is ingenuously introduced as the cause *of an action*. If my agent-self produces only a set of events, it may seem that I shall not have enough for my involvement in the action" (Williams 1995d, 71).

If there is only one species of genuine action, and it necessarily follows when the self or I causes it, by commanding or willing it, this can leave the impression that all those instances of human activity that do not involve this picture of willing are mere events. This, following Nietzsche-Williams, encourages a splitting or doubling in agency: there is an agent that produces mere events, and a self that causes actions. The "agent" being conceived in terms of a body or a kind of passive agency that merely responds to environmental stimuli, and the latter, as active, and reflective of the power of self-determination. The double counting is also taken to follow from the mode of causation being that of command: to obey a command can be conceived of as itself an action but, equally, to make a command is also conceived of as an action. It can seem like there are *two* actions in this picture of action.

As I understand it, the peculiar tendency to engage in double counting at the level of entities, and actions, makes the question of what the purpose of the picture might be more pressing. The unmediated relation between these special items, "a real action" and "a person or self or agent", offered by this picture of action serves, according to Williams, the purpose of supporting MS-blame (i.e., the morality system's *purified* conception of blame). The conceptual form of

blame is such that it requires “an occasion – an action – and a target – the person who did the action and who goes on from the action to meet the blame” (Williams 1995d, 72). This is straightforwardly the form of any relevant conception of blame whether it belongs to the morality system or not. But, on the morality system’s version of blame, the occasion, the transgression of a moral obligation, and the target, the will/self/person/agent, are needed in the isolated or pure forms found in this picture of action: the simple and pure conception of the will, and its unmediated, necessary relation to the special species of action, “real action”.

The fit between MS-blame and this picture of action in terms of the will, is owing to the demands of the morality system’s commitment to morality always being internally just. Recall, the system’s requirement that agents be blamed or held morally responsible “for no more and no less than what is in their power” (Williams 1995d, 72). If blame is going to be used, it must be done only in cases where an action was in an agent’s power. But, as we’ve seen, agents are subject to luck in all manner of ways, which is why determining what counts as in an agent’s power is “notoriously problematic”.

Conveniently, this picture of action represents the will not straightforwardly as the source of power in action but as a source of power in *what we do*, where “what we do” is pictured as free from empirical determination in exactly the way the morality system requires to maintain its requirement. This independence or isolation affords the space in which we can understand the will as a source of *complete* or *absolute* power in agency. Moreover, if the will/person/agent, as this picture of action supports, is to be understood as isolated from the contingencies of luck, and the actions it causes as its direct product, then the actions can thereby be taken as *necessarily* and *purely* reflecting the intentions or motivations of the agent/person. If this is the case, the necessary and pure connection this picture of action proffers between agent/person and action, provides secure grounds for correct moral judgement with respect to blame and moral responsibility as conceived in the morality system.

1.5.3 *Where does the picture of the simple and pure will come from?*

Now we can say what the *purpose* served by this picture of action as governed by pure will is: very basically, it secures grounds for correct moral judgement, blame, and moral responsibility. With this in place, I will turn to Williams' question of *where* this peculiar piece of our (moral) psychology comes from. Taking inspiration from Nietzsche's debunking genealogy of morality in terms of *ressentiment*, Williams outlines a speculative genealogy of the phenomenology of "focussed blame". As I interpret it, he does this in order to illuminate the basis of the link between MS-blame, and the system's investment in the picture of action in terms of the will. In what follows, I explain the genealogy, which leads into outlining another piece of our moral psychology, a fantasy, that Williams locates at the centre of MS-blame, and a desire for retributive punishment. Finally, I consider Williams' objections against this picture of action in terms of the will, i.e., the morality system's conception of the purely voluntary.

1.5.3.1 *Williams' genealogy of focussed blame*

Focussed blame or the focussed application of blame, the phenomenology of which is something like Nietzsche's *ressentiment*, is blame that is directed at an individual agent for particular acts or omissions.²⁰ The genealogy begins with a basic description of the dynamics or operations of focussed blame at the level of the experience of the victim:

If there is a victim with a complaint for a loss, there is an agent who is to blame, and an act of that agent which brought about the loss. The anger of the victim travels from the loss to the act to the agent; and compensation or recompense given by the agent will acknowledge both the loss and the fact that he was the cause of the loss (Williams 1995d, 73).

From the perspective of a victim who is angry at suffering a loss such that she has a complaint for the loss, there is an action that brought about the loss, and an agent who is blameworthy or responsible for performing the action that lead

²⁰ For Williams' other discussion of focussed blame see "Internal reasons and the obscurity of blame" (1995c, 40-41).

to the loss. The loss itself gives rise to the victim's anger, which (taking the form of focussed blame) moves to the action, and then to the agent who is seen to have caused the loss. Should the agent respond to the victim's blame with compensation or reparation, the character of it, from the perspective of the victim, should take the form of the twinned acknowledgements of the loss to the victim, and the fact that the agent was the *cause* of the loss.²¹

Williams, then, considers a further elaboration of the original scenario:

Suppose the agent brings about a harm to the victim, and does so intentionally and voluntarily; where 'intentionally and voluntarily' is not supposed to invoke the special mechanisms of the will, but merely means that agent knew what he was doing, wanted to do it, and was in a normal state of mind when he did it. Suppose that the agent is not disposed to give compensation or reparation, and that the victim has no power to extract any such thing from him. In refusing reparation, the agent refuses to acknowledge the victim or his loss; it is a peculiarly vivid demonstration of the victim's powerlessness (1995d, 73).

On one side, there is the harmed victim who, in their focussed blame, requires acknowledgement for the harm *intentionally* inflicted upon her by an agent. This expectation/need is amplified in the situation where the victim was harmed intentionally or voluntarily owing to the fact that this kind of action is typically taken to be, unequivocally, a display of ill-will towards the victim on the part of the agent.²² On the other side is the agent, their decidedly voluntary action, and its "success". If the agent is not disposed to give compensation or reparation for their action it stands as an agent's refusal to acknowledge the victim and their loss.

The victim, then, is left with a frustrated desire, possibly a need, for the acknowledgement of the agent who intentionally caused the harm. Williams also mentions the shame a victim may feel in this connection. In a situation such as this, the futility of the victim's blame with respect to moving the agent makes perspicuous the powerlessness of the victim to change the offending agent's

²¹ The suggestion here of the importance of the recognition of the significance of the loss to that particular victim in any given reparation or compensation will take on further relevance in Chapter 3 when I briefly sketch a model of reparation.

²² On the relevance of "ill-will" in a similar connection see Strawson (2008).

response to them, and, following Williams, conditions a basic feeling, something like *ressentiment*.

1.5.3.2 Williams on “the special fantasy of retrospective prevention”

According to Williams, the scenario of unacknowledged harm caused by voluntary or intentional action, and the powerlessness of the victim in light of that, also gives rise to an associated “special fantasy of retrospective prevention”: “As a victim, I have a fantasy of inserting into the agent an acknowledgement of me, to take the place of exactly the act that harmed me. I want to think that he might have acknowledged me, that he might have been prevented from harming me” (1995d, 73). The fantasy is retrospective in that it is directed to a point in time in the past, immediately before the agent took the action. The fantasy is concerned with prevention insofar as the victim imagines that, at this moment, before the action took place, she could have prevented the action that harmed her.

But, importantly, the prevention this fantasy is concerned with is *not* empirical, i.e., that in some empirical way I may have prevented the agent from doing the action. For example, the belief might form that, had I known the agent was going to take the action he took, I could have done *x* to prevent it. If this were the case, the fantasy would be structured as a regret that I, or anyone else, had actually prevented the agent from doing the action. As Williams appears to understand it, the fantasy is not about what I or anyone else might have actually done to change the circumstances or the way things went. In that case, what kind of prevention or change is the fantasy concerned with?

The fantasy is structured by the thought “that I, now, might change the agent from one who did not acknowledge me to one who did” (Williams 1995d, 73). What this fantasy requires, according to Williams, is the picture of the will as a simple and pure imperative kind of cause. Namely, it requires “simply the idea of the agent at the moment of the action, of the action that harmed me, and of the refusal of that action, all isolated from the network of circumstances in which his action was actually embedded” (Williams 1995d, 73). The fantasy involves the belief in a picture of the will, the source of power in action, as

isolated from the network of circumstances such that there are certain grounds (assuming certain conditions of agency) to believe that the agent could have, at the moment of action, refused to take the harming action.

Another way of putting this is to say that, if the will is conceived of as something free from empirical determination, then it operates in action outside the confines of circumstances, luck and character. This supports the victim's view that they could change the agent from one who did not acknowledge them, to one who would acknowledge them. I take it that the agent's character and their reasons for action can be put aside as irrelevant with respect to assessing the decision and action the agent took because there is a source of power in action (in the agent) that is free from these considerations. It secures the grounds for the victim's fantasy insofar as believing this picture of the will affords the possibility that, at the moment before action, the victim really could have inserted into the agent the acknowledgement to take the place of exactly the act that harmed her. Moreover, as I understand Williams, it is *this* fantasy, its operations and expectations that are reflected in MS-blame.

1.5.3.3 Williams on retributive punishment and the moral law

This basic feeling of *ressentiment* also "lays the foundation for the purest and simplest construction of punishment" (Williams 1995d, 73). What Williams has in mind is retributive punishment; and its power is again made vivid by the kind of example illustrated above. If the agent has no desire or interest in providing reparations we can easily imagine the victim desiring retribution to acquire some satisfaction or assuagement for their loss, and the lack of acknowledgement, by other means. Retribution, as I understand it, is a punishment inflicted on an agent as redress for an action they've performed that is considered morally wrong and/or a criminal act.

Insofar as the retribution is exclusively focused on the agent and the past action, it is, in its purest form, not concerned with actual reform, e.g., through "education, conversion or improvement" (Williams 1995d, 73-74). According to Williams, the more pure the retributive punishment, the more exclusive its focus is on the cause, which can be sufficient grounds for punishment.

Moreover, the punishment or hard treatment the agent is given, on this picture, can be deserved with “certainty”. What, then, is important for the grounds of assessment of desert is the familiar idea that it was in an agent’s power (i.e., it is just). Of course, if the picture of the will as simple, pure and an imperative kind of cause that brings its effect, an action, as if out of nothing, then there is a site of power in agency free from luck, and thereby, grounds for supporting the idea that punishment can be *intrinsically* deserved.

This construction of the pure will in action that supports the idea of intrinsically deserved retribution can operate in situations of actual punishment, e.g., a prison sentence. It is also, according to Williams, operating in our use and understanding of the morality system’s “purely moral” conceptions of blame and remorse. When operating in connection with these conceptions the construction involves, for Williams, a further abstraction: “it introduces not only retribution’s idea of retrospective causation, but morality’s idea of an authoritative but sanctionless law, of a judgement that carries no power besides that judgement itself” (1995d, 74). In other words, this picture of the will in action that the system encourages upholds the idea of intrinsically deserved punishment for moral wrongdoing. But, this is not restricted to actual punishments, and is further abstracted: the intrinsically deserved punishment happens in the face of the system’s authoritative but sanctionless abstraction, “the moral law”.

1.5.4 Williams’ criticisms of this conception of the will in action

Williams takes four diagnostic leads from Nietzsche on this authority bearing picture of action: (i) that the will is taken as something simple when it is not, (ii) that the will is taken as something pure, i.e., as free from empirical determination, (iii) that the will, as simple and pure, is a peculiar imperative kind of cause that effects a metaphysically distinct thing, “a real action”, and, finally, this picture of the will and its activity leads to (iv) a kind of double

counting involved at the level of entities and what's going on in action.²³ On the first and second diagnostic points, Williams claims that, maintaining the idea that the will is simple and pure rather than something complicated, forgets the difficulties and obscurities inherent in our attempts to describe the phenomenon of willing. In ignoring these descriptive difficulties and obscurities we forget the lack of transparency in our understanding of what is going on when an agent is willing. Attention to these difficulties and obscurities, Williams claims, need not lead to a radical scepticism regarding, specifically, the phenomenon of willing or the will, or more generally, the genuine presence of an agent in action (1995d, 74).²⁴

On another level, Williams is sceptical of this picture of action and its authority because it is neither coherent nor universal. The tendency of this picture of action to produce two entities, and two actions or activities, makes it incoherent. There is the commanding self and the obedient self, as well as an agent-cause, who brings about mere events, and an agent-self, who through willing brings about *real* actions. This encourages the idea that anything that merely happens (is not intended) is not an action and does not belong to or cannot be a reflection of the real agent-self/the will, and therefore, cannot be a proper object of moral assessment.

As I mentioned above, the confusion this double counting encourages can be read as a reflection of the morality system's deep investment in a picture of action that affords the possibility of complete power in what we do; as we see is required by the purified conception of blame this picture of action directly fits. The incoherence of the double counting in this picture of action can be understood as arising out of the needs or requirements of the system, which should be enough, according to Williams, to make us suspicious of *this* picture of voluntary action. "The fit between the special psychological conception [of action in terms of the will] and the demands of morality enables us to see that

²³ Kant is, of course, guilty of upholding the ideas that the will is something simple in the sense of it being "unified" and "pure". He also upholds the idea that the will is a peculiar, imperative kind of cause. But, interestingly, Kant cannot be said to be guilty of adhering to the idea of there being a commanding and an obedient self in action.

²⁴ That Williams genuinely takes this to be the case, I think, is evident in his non-purified conception of the voluntary (e.g., 2006a, 194; 1995b, 25; 1993a, 66), and his account of the will in connection with ancient Greek ideas about efforts of the will (1993a, 36-38).

this piece of psychology is itself a moral conception, and one that shares notably doubtful features of that particular morality itself” (Williams 1995d, 74).

Any objector to Williams’ claim may appeal to the idea that the picture is a universal one. But, as Williams points out, one need only look to the fact that other cultures (past and present) have, and do, uphold different pictures of action from this one to register that this picture is not universal. Williams uses the ancient Greeks to make his point. In *SN*, one of the central aims is to argue against readings of Greek, particularly Homeric, poetry and literature as lacking an adequate conception of voluntary action and responsibility, owing to the supposed absence of a concept of the (pure) will. The worry is then encouraged that, without the picture of the (pure) will, the Greeks did not have an adequate conception of action, specifically of the integrity of action, of the agent’s genuine presence in action (Williams 1995d, 74). But Williams sees this expectation as producing a misguided misinterpretation informed primarily by a crude Cartesian philosophy; a Cartesian misinterpretation of Homer (2006d, 66; 1993a, 21-49).²⁵

In several places Williams argues that, though the Greeks may not have used our concept of the will, we can nevertheless make significant sense of their picture(s) of action. Indeed, for Williams, we can *recover* from a serious and considered analysis of Greek texts, for example, what is basic to making sense of human action across historical and cultural divides (2006d, 65-66). And, in so doing, we are afforded a conceptual space into which we can move away from the illusions we find in the morality system’s picture of voluntary action, and towards a more “realistic” conception of action; a conception of action that perhaps, despite initial appearances, preserves the integrity of action, i.e., the agent’s genuine presence in it (Williams 1995d, 74).

1.6 What, then, is MS-remorse?

How does understanding Williams on the purified conception of voluntary action help to capture MS-remorse? Again, so far, I have uncovered

²⁵ This error has the same structure as the error Collingwood claimed of “the old gang of Oxford realists” who “would insist on translating some ancient Greek expression as ‘moral obligation’ and then point out that Aristotle, or whoever it was, had an inadequate theory of moral obligation” (Williams 2006c, 181).

that MS-remorse is the first-personal moral response to the voluntary breaking of moral obligations and, as such, the first-personal expression of moral responsibility to wrongdoing. The essential thought in MS-remorse is something like “all things considered what I did was wrong” or “I did what I ought not to have done”. Moreover, rational remorse is never a response to moral remainders or moral costs, nor to instances of involuntary action (e.g., unavoidable wrongdoing). What is a rational response to non-moral disagreeable aspects of right moral action, and instances of involuntary action, is MS-remorse’s non-moral counterpart, regret.

If I continue to assume the mirroring of MS-blame in MS-remorse, what I outlined above can be directly applied to MS-remorse. If, in connection with MS-remorse, I consider the simple and pure will, its connection with pure retributive punishment, and the idea of a judgement that carries no power besides that judgement itself, I get a picture of retributive *self*-punishment that is peculiarly internal and abstracted. I take it that these ideas encourage a picture of moral response that sees an agent focusing on punishing *herself* for her failure *with respect to* that other part of herself, her pure will, in light of her anticipated condemnation from an absolutely authoritative, abstract moral system (e.g., a moral law conception of morality). On my diagnosis, when pure retributive punishment connects with the purified picture of voluntary action it encourages a picture of self-punishment in remorse (or guilt) that, in positing two entities, can conjure a peculiar form of self-involvement directed by one’s sense of having failed an abstract authoritative “moral law”.

My point is not about whether this kind of self-punishment in remorse (or guilt) tends towards excessive forms of punishment (e.g., self-torment). Rather, my point is that it encourages a picture of remorse that is excessively self-involved and attendant to an abstraction. Williams would, I think, following Melanie Klein, call this “persecutory guilt” as opposed to “reparative guilt” (1973b, 222). In being self-involved and responsive to an abstraction, what I think puts this picture of remorse or guilt under immediate critical, and moral, pressure, is the impossibility of anything other than unproductive self-punishment.

As Williams puts it when he discusses moral sincerity in connection with remorse and guilt, “Those uncreative aspects of guilt...[may] encourage philosophical scepticism about the relevance of the emotion; unproductive self-punishment may be seen as precisely *not* an expression of those principles that ought to have issued in action, but rather as a misdirected substitute for action” (1973b, 222). I take it that what renders the self-punishment unproductive is the self-involvement encouraged by the double counting of entities found in the purified conception of voluntary action. When coupled with the authoritative moral law, it makes it difficult to imagine how this picture of guilt could ever be productive of genuinely other-regarding reparative action; something I take to be a cardinal feature of any conception of remorse or guilt. But, following Williams, all is not lost. For, as Klein’s distinction suggests, this is not the only picture of remorse (or guilt) we have available to us.

Finally, I think we are in a position to return to my opening question: how does MS-remorse capture a concern for people directly affected by one’s action? The short answer is: it doesn’t, or at least not directly. The long answer, to the effect that MS-remorse captures rather a concern for principles violated by one’s action, will be made out in the next chapter where I consider Gabriele Taylor’s separate accounts of guilt and remorse. As we will see, Taylor’s account of guilt is a clear example of the morality system’s conception (guilt as “MS-remorse”), and though her separate account of remorse partakes of features of MS-remorse, surprisingly it is rather her account of guilt that more loosely resembles what we might expect from an a first-personal expression of responsibility for wrongdoing: namely concern for people directly affected by one’s action, and the attempt, where possible, to make reparations.

CHAPTER 2

TAYLOR ON GUILT AND REMORSE

2.1 Introduction

At the time of Gabriele Taylor's *Pride, Shame and Guilt: Emotions of Self-Assessment* (1985), there was an upsurge in philosophical interest in the emotions, and, a related interest in emotions in moral psychology. Amélie Rorty (1980), Ronald de Sousa (1987), Patricia Greenspan (1988), and Martha Nussbaum (1990) are just some of the philosophers who have played a critical role in bringing the emotions back into philosophical debate and, more specifically, into moral psychology. Many of these philosophers, and others, recognise Taylor's seminal contribution to the foregrounding of moral emotions. This is reflected in the frequent inclusion of Taylor's ideas on guilt and shame in the literature as either grounding the contemporary debates and/or as offering legitimate anchoring or starting points. Yet, it is rare that her embedded account of guilt, and its other companion moral emotion, remorse, receive the focused critical attention that an account with this general level of influence merits. Admittedly, this neglect is surely, in part, due to the difficulties of untangling the different aspects of her view. The critical reconstruction of Taylor's accounts of guilt and remorse that I give in this chapter, however, is not principally aimed at making good their relative neglect in the debate. Rather, I reconstruct Taylor because it is revealing in relation to the prospective genealogy of remorse that I aim to offer in the thesis overall. Taylor is unusual in offering an account of remorse at all, but more importantly her account betrays the ways in which remorse, as we know it, has been unduly constrained by the morality system (as it finds expression in moral philosophy).

Why examine both guilt and remorse? Guilt, as Taylor rightly assumes, is absolutely central to moral psychology. Given this, she understands that anyone with an interest in understanding guilt or remorse needs to examine how these moral emotions are related. As I understand it, we can ask whether guilt and remorse are the same thing, whether they are differently focused or inflected, or perhaps whether they are properly distinct. Taylor falls into the third option,

owing to her view that guilt, unlike remorse, is an emotion of self-assessment.²⁶ Though distinct, Taylor claims that guilt and remorse are proximate insofar as they can be felt about the same event.

The central and abiding insight of Taylor's account is her claim that guilt is an emotion of self-assessment. But, as we will see, despite this correct starting point, Taylor's account of guilt steadily goes wrong. I argue that she makes two critical errors. The first error concerns her conception of self-assessment, which produces an unduly egocentric or self-involved conception of guilt. The second error concerns the influence of the morality system's conception of morality as necessarily binding and absolute so that any moral wrongdoing is conceived as a transgression of some absolute duty or principle. What results is an account of guilt that upholds wrongdoing as a transgression of (something equivalent to) "the moral law", and is essentially concerned with the self, and how one has damaged oneself in transgressing the absolute strictures of morality. Following a line of criticism pursued by Bernard Williams, and Raimond Gaita, I will argue that Taylor's adherence to the morality system's absolutist conception of moral authority, combined with her egocentric focus in self-assessment, together lead to a loss of moral content in her conception of wrongdoing. On her account of guilt, what surely should be the focus of any account of guilt or remorse has gone missing: the harm done to others.

Given Taylor's commitment to an utterly self-involved notion of self-assessment, when she claims that remorse is not an emotion of self-assessment, she takes it to follow that remorse is a distinctively other-regarding moral emotion. This, I shall argue, is owing to Taylor's systemic dualistic assumption: that for any given moral emotion it is either strictly self-regarding or strictly other-regarding. She therefore neglects the possibility of an emotion being neither strictly self-regarding nor strictly other-regarding, but rather *relational*. In addition to this error, Taylor's account of remorse goes wrong in at least two ways. The first concerns her insistence that remorse is a distinct emotion from regret, instead of recognising that remorse is better understood as a sub-

²⁶ Confusingly, at the start of the text Taylor claims that remorse is also an emotion of self-assessment (1985, 1). But, she goes on to claim that, "remorse is not an emotion of self-assessment" (1985, 99; 103). Throughout this chapter, I follow her claim that remorse is not an emotion of self-assessment.

category of regret. What results is an excessively polarised conception of remorse on the one hand, and regret on the other, which produces a picture of regret as a mere artefact of what she claims regarding remorse. The second respect in which I shall argue that Taylor goes wrong is in failing to acknowledge that her account of remorse in fact shares most of its main features with her account of guilt. Taken together, the false dualism and the very substantial overlap between the accounts of regret and remorse lead one to question whether guilt and remorse are really as distinct as Taylor makes out. Ultimately, what my critique of Taylor's account aims to deliver is a vindication of the natural thought that either there is simply no difference between guilt and remorse properly construed, or, alternatively, remorse is a special form of the moral emotion of guilt, made distinctive by its particular focus on the harm done.

The chapter has four sections. The first section presents Taylor's project, including her problematic concept of self-assessment, and her theoretical framework. Sections two and three respectively reconstruct and criticise Taylor's accounts of guilt and remorse. Given the sometimes entwined and confusing character of the two accounts, I have extracted and ordered what I take to be the supposedly independent essential features of each. This model of reconstruction provides the grounding for separate critiques of Taylor's accounts of guilt, and remorse, at the same time as grounding the final section's critique of the weakness of Taylor's attempt to distinguish these emotions.

2.2 Taylor's project

2.2.1 Taylor on self-assessment

Pride, Shame and Guilt: Emotions of Self-Assessment is broadly concerned with what Taylor takes to be the unifying feature of pride, shame and guilt: they are emotions of self-assessment. Taylor's claim is that in experiencing pride, shame or guilt the agent takes herself to have deviated from some norm, moral or otherwise, and consequently has "altered her standing in the world" (1985, 1). In the case of pride, the agent's standing is typically raised. By contrast, shame and guilt are in different ways concerned with a lowering of moral

standing. The general structure of these emotions of self-assessment, for Taylor, takes the self as its object, and the assessment of that self as constituted by what the agent believes. Moreover, the change in one's standing, regardless of whether it goes up or down, is reflected in the view the agent takes of himself: "Starting from a set of beliefs or assumptions about himself, his conception of some event or state of affairs is such that he has to formulate beliefs about himself which conflict with the ones held initially; so he has to alter his view of himself" (Taylor 1985, 15-16). Here, I am interpreting Taylor to take the alteration in one's beliefs and perspective of oneself to be relative to one's "whole self" as opposed to aspects or parts of oneself.²⁷

Again, Taylor's insight that guilt is an emotion of self-assessment is what her account gets right. It introduces the idea that guilt is an emotion that involves not merely an assessment of an act but also, typically, an evaluation of one's self in relation to that act. I feel guilty, let us imagine, because I betrayed my friend's trust by revealing her deepest personal secret to a mutual group of acquaintances in order to be seen by them as interesting. On Taylor's account, we can say that my belief that I am a steadfast and trustworthy friend comes into conflict with my betrayal of my friend. Compounding this further is the fact that I betrayed my friend for a shallow reason. I can no longer sustain the belief that I am a steadfast and trustworthy friend nor the belief that I am not influenced by appearances, as I previously took myself to be. On Taylor's account, I must alter my view of myself as no longer someone who is such a friend (though it need not follow that I now view myself as disloyal and untrustworthy).

Although surely correct in her central claim, there are two ways in which Taylor is too extreme in her account of self-assessment. The first relates to the character of the alteration in self-assessment. She claims it always involves an alteration of beliefs about oneself of a kind that constitutes an alteration in one's

²⁷ This interpretation is supported by Taylor's distinction between primary and secondary deviance in self-assessment. Primary deviance concerns the acceptance of having done something wrong but taking it to be alien to "what one really is", e.g., I cheated on my test but I am not really a person who cheats (1985, 90). Secondary deviance takes the wrongdoing to be expressive of "what one really is", e.g., I committed a burglary and now see myself as a burglar (1985, 90). That the distinction marks the difference in how one takes the wrongdoing to relate to "what one really is" supports the interpretation of Taylor that the change in belief and perspective in self-assessment is relative to one's "whole self".

view of oneself. But this is not so. On the contrary, feelings of shame or guilt are typically associated with a belief of the form “I have let myself down”, and such a belief actively presupposes a certain stability in one’s view of oneself. Taking the example above, I may believe that I betrayed my friend for a superficial reason, and that I have let her, and myself, down in doing so. But this need not lead me to relinquish my fundamental belief in myself as a steadfast and loyal friend. The negative moral emotions of shame or guilt may operate rather as an acknowledgement that my behaviour reflects well or badly on me precisely in relation to that stable view of myself. In response to my feelings of shame or guilt, I may (contra Taylor) actively affirm my existing view of myself by resolving not to act that way again. Taylor effectively neglects the possibility that shame and guilt may often, indeed normally, conserve one’s view of oneself, and give rise to an attitudinal change of a more localised kind, such as the resolve to act in a manner more consonant with one’s view of oneself.

The second way Taylor’s conception is too extreme is in her interpretation of the idea that self-assessment takes the self as its object. For Taylor, it means that the self is the *exclusive* object of concern in any emotion of self-assessment. As I will show, this interpretation arises in connection with Taylor’s falsely dualistic assumption that an emotion can only be either self-regarding or other-regarding, never both at once. This assumption in her account of guilt results in a picture where the assessment of the self occurs not in direct relation to others, but in direct relation to an authoritative abstraction, e.g., “the moral law”. But, as the aforementioned example suggests, my assessment of myself can occur with respect to how I treated a particular person (e.g., my friend), and not strictly in terms independent of others (e.g., with reference to “the moral law”). If I want to assess myself as a friend, I must focus not only on myself but also on my friends—how I have behaved in relation to them. In fact, this relational mode may often apply even for self-assessments regarding roles that are less obviously relational in kind. If, for instance, I want to know whether I am a good cook, I may look to my partner’s responses to my cooking for an assessment. In short, there is simply no reason to think that self-assessment is not commonly relational in its focus. That Taylor fails to see this explains why her conception of self-assessment becomes unduly egocentric in

form. Before elaborating how this self-focus affects her account of guilt, I shall present Taylor's cognitivist view of the emotions.

2.2.2 Taylor's framework: a cognitivist view of the emotions

Taylor takes a cognitivist view of the emotions: emotions as constituted by beliefs. Specifically, she maintains that our emotional experiences are constituted by what she calls "identificatory" and "explanatory" beliefs (Taylor 1985, 2-3). These two types of beliefs are beliefs about the object of the emotion, without which there would be no emotional experience. Identificatory beliefs are those belief(s) that identify the emotion as being what it is, and thereby differentiate it as say fear, or envy, as opposed to grief or anger. Explanatory beliefs are those beliefs that ground or make rationally intelligible the identificatory beliefs. Typically, for example, we feel fear in a situation we believe to be dangerous or harmful (identificatory belief). What will explain why I take the situation to be dangerous or harmful will be some feature(s) of the situation in virtue of which I believe it to be dangerous (explanatory belief). For example, if Janet is afraid of the scorpion, then there will be some feature Janet takes it to have that makes it dangerous, e.g., that it is poisonous. Janet is afraid of the scorpion *because* its sting is poisonous and could thereby harm her. In this way, the explanatory belief "its sting is poisonous" is "causally responsible" for the fear being what it is. Janet would not take the scorpion to be dangerous were it not for her belief that the scorpion is poisonous.

However, according to Taylor, there are ways in which the beliefs may fail to provide a rational explanation for a given emotional state of a particular person. Taylor acknowledges that the success or failure of these beliefs providing explanation will depend on such considerations as: (a) whether the beliefs are well-founded or rationally justified themselves, (b) whether the beliefs are justifiable in the light of the actual circumstances, and (c) even if the beliefs are justified and correct, whether the feature(s) make the situation what the person takes it to be (i.e., dangerous or harmful).

These dimensions of assessment can be performed by the person experiencing the emotion and/or by an observer interested in explaining why

the person feels the way she does in a given situation. Sensible to this possibility, Taylor recognises that both the person, and any second or third party, may draw on or refer to other beliefs the person has held in previous situations. This may facilitate the intelligibility of the current beliefs that constitute her reasons for her emotional experience (Taylor 1985, 14). In acknowledging the relevance of these diverse considerations in any given explanation of a person's emotional state, Taylor is distancing herself from a strain of cognitivist theory of emotions that sees the constraints of explanation as strictly drawn by the "fully rational".²⁸ In this connection, Taylor considers Donald Davidson's account of rational intelligibility in emotion (1985, 5-14).²⁹ Davidson, according to Taylor, takes it that the rational explanation of an emotion requires a belief that is universal in form, such that "the required explanation can take the form of a syllogism, so that the reasoning which leads the person to the conclusion that he is praiseworthy or guilty in a certain respect is deductive" (Taylor 1985, 6). Despite acknowledging the plausibility of Davidson's view, Taylor objects claiming that not only is it "too neat and simple", but that "a detailed explanation in terms of particular beliefs may make the person's belief that *p* [e.g., "I am praiseworthy", "I am scared"] intelligible in a much more satisfactory" (Taylor 1985, 8).

In advancing the idea that the intelligibility of emotions need not depend on a notion of the "fully rational" in the manner exemplified by Davidson, Taylor diverges from that kind of reductivist strain in cognitivist theories of emotion. Yet, she is subject to other familiar criticisms against cognitivist theories. In *The Emotions: A Philosophical Exploration*, Peter Goldie is variously critical of the kind of cognitivist view Taylor presents. Goldie takes cognitivist positions like Taylor's to reduce emotional experiences to feelingless beliefs, and thereby, take a cardinal feature—feelings—out of emotion. This, I take it, tends to arise on such accounts because of the priority they confer to one's judgement, ignoring the fact that feelings are often prior to, or in conjunction with, the relevant judgement.

²⁸ See Deigh (1994) for a full critique of such views.

²⁹ Taylor uses Davidson's paper "Hume's Cognitive Theory of Pride" (1976).

Goldie is also critical of what he takes to be the reduction of the explanation of, or “making sense of”, emotions to a thin conception of intelligibility. That is, explaining emotions strictly in terms of beliefs that are generally considered to ground it. As I interpret it, this criticism aims to capture the idea that the potential complexities of many emotional experiences are, on such cognitivist accounts, ignored as at most explanatorily marginal, if not irrelevant, with respect to the intelligibility of emotions. According to Goldie, there are many critical dimensions integral to the making sense of emotions alongside intelligibility. These critical dimensions include whether the beliefs involved are rationally arrived at; context-sensitive appropriateness and proportionality; and a person’s mood, character and their other emotions (Goldie 2000, 23). Clearly, from the above, I take Taylor to be sensible to some of these other dimensions of assessment but she, unlike Goldie, does not place them on equal grounds of relevance in explanation owing to the character of the priority she gives to intelligibility in terms of beliefs.

The final aspect of Taylor’s cognitivist view of the emotions to consider is that identificatory beliefs have a *general* quality. Take the case of the identificatory belief of fear (i.e. the belief that constitutes it as a case of fear): I believe the object or situation confronting me to be dangerous or harmful. It is a necessary condition for the experience of fear that the person believes the object to be dangerous or harmful whether or not their belief is justified. The identificatory belief, therefore, presents the basic structure of the emotional experience of fear, i.e., what is common to all experiences of fear irrespective of whether it turns out to be rational or justified. And this is what Taylor aims to uncover in the case of guilt: the identificatory belief that will be common to all experiences of guilt regardless of whether they turn out to be irrational. In light of her interest in uncovering the identificatory belief of guilt, Taylor focuses on the immediate experience of feeling guilt. She, therefore, includes in her account characteristic instances of irrational guilt where for example, an agent has yet to examine the rationality of their belief regarding a particular moral obligation or principle, and cases where guilt feelings persist despite a rational change in one’s belief. A familiar example of the latter is someone who now eschews the

religious beliefs he grew up with, but continues to feel guilt for not attending church on Sundays (Taylor 1985, 86).

As I will explain below, through the examination of the essential and non-essential features of guilt, Taylor concludes that the identificatory belief constitutive of guilt is that one has “done something forbidden and that in doing what is forbidden she has disfigured and so harmed herself” (1985, 103). It is worth noting that the identificatory belief, in its emphasis on the person’s relation to their “disfigured and harmed self” in guilt, reflects Taylor’s view that guilt, as an emotion of self-assessment, takes the self as its primary object. By contrast, in her discussion of remorse, Taylor takes the identificatory belief to be the following: “when feeling remorse the agent believes that she has done harm which she ought to try and repair” (1985, 104). The harm, according to Taylor, is not the disfigurement of the self in guilt but the harm done to others. At the end of this chapter, I will argue that the difference in these identificatory beliefs is not as marked as Taylor claims. This makes available what I take to be a natural thought Taylor works hard to avoid: that guilt and remorse are not in fact distinct, or, are not as distinct as she takes them to be. With Taylor’s cognitivist framework in place, I now turn to reconstructing her account of guilt in terms of its essential features.

2.3 Taylor’s account of guilt

As I argued in the previous chapter, the morality system upholds a conception of morality as always, and necessarily, binding. Often taking the form of an abstract “moral law”, under such a conception, any moral wrongdoing is understood as a transgression against the moral law in terms of the violation of one of its moral obligations or principles. Taylor’s account of guilt exemplifies the influence of the morality system in exactly this way for, as I will show in the next section, she takes the fundamental occasion of guilt to be the transgression of moral principles or obligations one regards as absolutely authoritative and binding. I argue that her commitment to the morality system’s notion of morality and wrongdoing, in combination with her commitment to an egocentric focus, produces a conception of guilt that loses the moral substance

of the wrongdoing: that is, the harm done by the agent, as suffered by others. These issues appear throughout Taylor's account of the essential features of her conception of guilt, and I shall discuss them in the following terms: acceptance of authority, the essential thought in guilt, essential and non-essential harm, causal responsibility, and repayment. In what follows, I disentangle what I take these features to be, and argue for how they ultimately compound the problems of her account.

2.3.1 *Acceptance of authority and Taylor's conception of moral authority*

Invoking guilt's legal associations, Taylor distinguishes the possibility of *being* guilty from *feeling* guilty. Though primarily interested in what is involved in *feeling* guilt, she uses *being* guilty to illuminate a necessary condition for feeling guilty. If a person breaks a publicly recognised law, she can be declared guilty of the violation, and is thereby liable to punishment. In this case, *that* she has violated the law is sufficient for her being guilty. But, the accused may not feel guilty because she thinks the law unjust or, perhaps, she doesn't recognise the authority of the law. According to Taylor, in order for a person to feel guilt,

[H]e must accept not only that he has done something which is forbidden, he must accept also that it is forbidden, and thereby accept the authority of whoever or whatever forbids it. The person who accepts the authority does not merely recognize its power and so thinks it simply prudent to obey its commands; he also accepts its verdicts as correct and binding. What the authority pronounces to be wrong must not be done (1985, 85).

On this account, it follows that it is a necessary condition for feeling guilt that one accept what one did as something that is "forbidden". Forbidden according to whom? The short answer is "an authority which issues commands". But, it is not simply that one accepts the authority's commands for prudential reasons or out of fear of punishment. Rather, one must also accept the authority's commands as correct and binding. The use of "forbidden", and the idea that "what the authority pronounces to be wrong *must not* be done", suggests the morality system's influential idea that in order to see something as the wrong

thing to do or have done, one must take the issuing authority's commands to be *absolutely* binding.

According to Taylor, the authority that issues these commands and prohibitions becomes "the voice of conscience" (1985, 85). She focuses on explaining the nature of the authority of the voice of conscience. But, the very reason Taylor needs to explain the nature of the authority of the voice of conscience is owing to the unexplained shift in her account from an external to an internal authority figure in the form of one's conscience. If there is an external authority figure threatening punishment or retribution for the breaking of its commands, I take it that it is easier to understand the nature of its power and authority (e.g., the threat of physical punishment or withdrawal of approval). If internalised, however, it is simply unclear what gives the voice of conscience an equivalent force, and a sense of bindingness. Taylor, therefore, needs an account of how the binding authority of an external figure is transfigured into one's conscience. Taylor's weak solution is to claim that as "an authority whose commands have to be obeyed has the status of a god" so the voice of conscience should be thought of as reflecting the edicts of some god (1985, 85). The point of the "god" metaphor, I take it, is that the voice of conscience, as an inner voice guiding the rightness or wrongness of one's behaviour, takes on the character of representing the commands of an absolute, objective moral authority.

2.3.1.1 Criticising the concomitant picture of guilt

Although unclear in her account, surely the philosophical influence being invoked by Taylor is Kantian: the thought of conscience as taking the form of an inner "Judge" who represents "the moral law". Bernard Williams aptly describes a picture of guilt that tends to emerge from this way of thinking about moral authority in the following way: "the point is supposedly reached where there is no distance at all between subject and internalised figure, and guilt is pictured as an emotion experienced in the face of an abstraction, "the moral law", which has become part of the subject himself", where guilt comes "to be represented simply as the attitude of respect for an abstract law" (Williams 1993a, 219-220;

222). Leaving aside the obvious similarity between this picture of guilt and my picture of MS-remorse, Taylor is basically advancing this picture of guilt. For the present, the immediate consideration is that the internalised figure in Taylor's model is the conscience, which, in guilt, responds to what the objective moral authority deems "forbidden".

Taking conscience as representative of an absolute, objective moral authority, explains, according to Taylor, what makes guilt a *moral* response. Although Taylor has no interest in claiming that there is *a* correct moral system, she is adhering to the idea, familiar from the elaboration of the morality system, that what makes something moral is having the character of being absolutely authoritative and binding. The morality system is thus revealed as a generic influence on Taylor's conception of morality: she adopts its absolutist conception of morality, and as a result her account of the experience of guilt is one of having transgressed a god-like authority.

The objection against this picture, which I will pursue here and in later chapters, is the one identified by Williams, and Raimond Gaita. Though their diagnoses are different, both Williams and Gaita take the kind of picture of guilt Taylor is advancing to be one that loses sight of those harmed by what one has done. Namely, it suffers from a loss of moral content, or the substance of any particular wrongdoing. Williams elaborates the objection in the following way,

When the conception of guilt is refined beyond a certain point and forgets its primitive materials of anger and fear, guilt comes to be represented simply as the attitude of respect for an abstract law, and it then no longer has any special connection with victims. The victims may re-enter, of course, in an account of what the subject has done in violation of that law, but that gives them, or thoughts of them, no more intimate connection to guilt than they have to shame (1993a, 223).

As we shall see below, on Taylor's account, those harmed by what one has done may re-enter in one of two ways. The first is the manner Williams identifies in the passage: in account of what the subject has done in violation of an abstract moral law. The second, as I will discuss in the final section of this chapter, is by way of the possibility of experiencing guilt, and other-regarding remorse over the same event.

2.3.2 Self-regarding guilt and the essential thought in guilt

As we saw in the previous section, accepting something as “forbidden” requires that one view the issuing authority as absolutely authoritative such that one takes its commands to be binding no matter what. Wrongdoing, therefore, involves the transgression of those commands or demands issued by the relevant moral authority. Keeping this in mind, we now turn to Taylor’s somewhat convoluted argument for what she takes to be the identifying belief of guilt: “what I am doing is forbidden”.

To arrive at this essential thought of guilt, Taylor builds her account of self-regarding guilt by setting up, and arguing against, a conception of guilt that focuses on harm done to others, what I will call “other-regarding guilt”. Taylor mistakenly takes John Rawls’ account as exemplifying other-regarding guilt. She claims that, for Rawls, guilt is felt in response to the “transgression of a principle of right”, such that “the wrong I think I have done must be thought of as some harm to another” (Taylor 1985, 86-87).³⁰ Taylor’s mistaken interpretation follows from ignoring the significance of Rawls’ distinction between guilt found in the morality of association (“association guilt”) and, at the “highest” level of moral developmental, guilt in the morality of principles (“authoritative guilt” or “guilt in the strict sense”) (1999, 412; 415-416). The “more mature” “authoritative guilt” is directed at the given principle of right, where the shared conception of “right” is authoritative independently of the contingencies of relations of association (i.e., a constructivist conception of moral law) (Rawls 1999, 415). When Taylor interprets Rawls’ use of “transgression of a principle of right” as involving taking the wrongdoing as in relation to harm to another, I take Taylor to be conflating Rawls’ “authoritative guilt” with his “association guilt”. In doing so, Taylor thus ignores that Rawls’ conception of wrongdoing in authoritative guilt is the same as the one she uses from the morality system. For Rawls’ authoritative guilt, just like Taylor’s conception, is a version of the picture of guilt as an emotion experienced in the face of an abstraction.

³⁰ See Deigh (1988) who also claims that Taylor misinterprets Rawls, and that her position is not as different as she makes out (392).

Although Taylor is mistaken about Rawls, for ease of explanation, I will follow her caricature of his account as a representative of an other-regarding conception of guilt. The first line of criticism Taylor pursues against this conception focuses on the descriptive suitability of “transgression of a principle of right” in all the relevant cases of guilt.³¹ She questions its suitability on two counts. The first concerns the explanatory reach of transgression of a principle of right. Taylor gives the example of feeling guilty if I don’t have the patience to listen to a person’s tale of woe. Conceiving of such a case in terms of a “transgression of a principle of right” would be, according to Taylor, a stretch. But, Taylor’s commitment to “the forbidden” makes her account subject to her own objection. Would the person’s feelings of guilt at not listening to another’s tale of woe be best characterised as her taking it to be “forbidden” not to listen? Would the person who feels guilty for indulging in watching television instead of reading think of it in terms of “forbiddenness”? Surely these descriptions also suffer from being too strong. I think that if Taylor were to disconnect her conception of authority from a notion of absolute authority, she might use a less formidable term to characterise what is correct about her thought: that, in experiencing guilt, we take ourselves to have done something wrong or morally bad.

The second objection regarding the suitability of “transgression of a principle of right” concerns a supposed implication associated with “rights”. Taylor claims that the reference to “rights” implies an “emphasis on persons other than the agent himself...that the agent’s thoughts are primarily on the rights of others” (1985, 87). As I interpret this, Taylor takes the use of “rights” to mistakenly place the emphasis in thought in guilt on others. The supposed implication continues, that if an agent thinks in terms of rights her perspective of the situation will be other-regarding. But why should thinking in terms of

³¹ How, and how far, does one draw the line with respect to the cases to include as determining the philosophical structure of a given emotion? There is an important methodological question here concerning whether seeking necessary and sufficient conditions with such a diverse range of cases has the effect of thinning or even distorting the theory’s power to explain the more standard cases. It may be that a different philosophical method is better suited to emotions like guilt, one that is concerned with modelling paradigmatic cases of an emotion in such a way as to be able to accommodate exceptional examples or pathologies of it without thinning or distorting the explanatory structure. See Fricker (2014) and Chappell (2014) for two different but related suggestions for alternative philosophical methods regarding how to conceptualise some moral experiences.

rights entail a focus on others in thought and perspective? Rights can be thought to be inscribed in the moral law such that, it may be that the principle one violated may be a principle of right, but, what one's thought focuses on in guilt is that it was a transgression of a principle upheld by the moral law. What makes the neglect of this possibility even more striking is that this is the picture of guilt Taylor is ultimately advancing.

It is mistaken, according to Taylor, to think that thought in guilt focuses on rights, and on persons other than the agent herself because, "it is true that the person who feels guilty thinks in terms of duties not performed and obligations not fulfilled" (1985, 87). Following from this "truth", Taylor simply asserts that thought in guilt is in terms of obligations and duties, and not rights (1985, 88). Moreover, she assumes that the perspective associated with thinking in terms of obligations is essentially self-regarding. By way of assertion and dualistic contrast, Taylor offers a picture of guilt that takes the thought in guilt to be focused on one's duties and obligations, which she takes to entail an essentially self-regarding perspective.

The following picture emerges in Taylor's account: there is self-regarding guilt, which focuses on one's duties and obligations, and there is the mistaken other-regarding guilt, which focuses on rights. In terms of obligations and rights, Taylor is over-extending the dualism because, as she acknowledges, obligations can imply rights and rights can imply obligations (1985, 88). However, she maintains a stronger dualism with regards to perspective. She writes that an agent may "Look at it from the point of view of how it most importantly concerns himself, or how it most importantly concerns the other" (Taylor 1985, 88). It seems correct to say that the perspective of how the wrongdoing most importantly concerns oneself will be incompatible with the perspective that considers how the wrongdoing most importantly concerns the other. However, what is doing the work is not a fundamental dualism in the possibility of perspective but the priority related to importance. If I consider the wrongdoing in terms of how it most importantly concerns myself, then this implies a priority of one's self over and above others. Indeed, *this* self-involved point of view is incompatible with recognising how one's wrongdoing most importantly concerns others. As I will explore in the Chapter 3, we can take a

perspective that registers concern for one's self without giving priority to one's self. This is a clear indication that Taylor is assuming that what it means for a perspective to be self-regarding or self-concerned at all is for it to be completely self-involved. It follows, then, that, for Taylor, the perspective in guilt is egoistic or highly self-involved. This reflects her commitment to her starting point: that guilt, as an emotion of self-assessment, takes the self as its object. What this therefore means, on Taylor's account, is that one's attention and concern in guilt is *necessarily*, and *exclusively*, regarding one's self.

The second line of criticism Taylor pursues against the conception of other-regarding guilt she has created is to point to the fact that, in feeling guilt, it is not necessarily the case that the character of the wrongdoing involves harm to another. Taylor focuses on examples where the wrongdoing harms oneself, e.g., in wasting one's time or abilities. Her example is the person who indulges in watching a silly television serial instead of reading literature to improve her mind (Taylor 1985, 88).³² She acknowledges that the reason for her guilt could be construed as indirectly harming those with whom she comes into contact in the present, and possibly, the future. But, Taylor replies, she may feel guilty because she believes indulging in something at the expense of self-improvement is wrong *on principle*. This shows, according to Taylor, that thought in guilt can be focused exclusively on harm to oneself. In light of this, Taylor claims we can say that the thought of harming another is *not* necessary in guilt. Given this, Taylor argues that the content of the wrongdoing need not be other-involving in any way, leading her to conclude that, "What is important for guilt is just that some form of action or abstention should present itself as obligatory to the agent, but the content of the demand is not restricted" (Taylor 1985, 89). This is a clear expression of the formalism that is part and parcel of the absolutist conception of moral authority characteristic of the morality system. What matters for guilt, as a moral response to wrongdoing, is that one regards the given demand as obligatory, i.e., as absolutely binding.

Taylor takes, as her argumentative target, a conception of guilt that maintains that one's thought and perspective in guilt are essentially other-

³² This example is illustrative of Taylor's broad view of morality, which includes socially reinforced rules and obligations as well as "personal morality", i.e., "a persons' own view of how he ought to live and what he ought to be" (1985, 76-77).

regarding. She argues against this view claiming that the guilty think in terms of their obligations and duties, and thereby necessarily take a self-regarding perspective. Taylor also criticises other-regarding guilt because it cannot capture the full range of guilt experiences; guilt can be felt in response to instances of harm to the self, e.g., breaking one's diet or not fulfilling one's potential. In other words, others do not enter into this kind of guilt experience at all. From her objections, Taylor concludes that the thought "I am harming another", and the taking of a perspective in guilt that focuses on another, is not necessary for the experience of guilt. I could think "I am harming another" in my guilt but, following the formalism of the morality system, I would think this because the content of the principle or command I have transgressed relates to others, and not because it is essential to my experience of guilt. Accordingly, Taylor concludes that, given that all experiences of guilt involve the transgression or violation of a moral principle or obligation (regardless of the nature of content), what must be common to all experiences of guilt is the essential thought "what I am doing is forbidden". When taken together with the acceptance of authority, we get: "what I am doing is forbidden", where "what is forbidden" is understood as whatever moral principles or obligations a given person upholds as authoritative and binding. Following from Taylor's picture of self-regarding guilt, I can now turn to the third essential feature of her conception of guilt: essential harm.

2.3.3 Essential harm and non-essential harm

Taken together Taylor's emphasis on one's conscience as inner judge, the absolute authority of the forbidden, the formalism of the morality system, and the self-involved perspective, it is clear that her conception of guilt is a representative of guilt pictured as experienced "in the face of an abstraction, the moral law, which has become part of the subject himself" (Williams 1993a, 219-220). Taylor is, therefore, subject to the objection that her account ignores or loses sight of the moral content of wrongdoing. It may seem that introducing a distinction between essential and non-essential harm could help Taylor's case. But, as my analysis will show, the distinction distances Taylor even further from

the immanent moral content of wrongdoing. Taylor describes her distinction in the following way:

The first, essential, type of harm is the direct consequence of the deed being forbidden; the second, non-essential type of harm is contingent on the nature of whatever it may be that agent sees as forbidden. That may be, for instance, violating another's rights, or it may be neglecting her talents. In the latter case, therefore, the person is harmed under two descriptions, that she has done what is wrong, and that she had not developed her talents. In this case doing the latter also happens to constitute the former, but the specific way in which she thinks she has harmed herself is distinguishable from the harm that is the stain of guilt (1985, 103).

The essential harm in guilt is the harm the agent brings upon herself simply in virtue of doing what she takes to be forbidden; *that* it is forbidden causes harm to the self. What makes this type of harm essential to guilt is its independent connection with "the forbidden". The idea seems to be that one's moral standing is lessened by being a wrongdoer *in the eyes of morality*. When a person does something she takes to be forbidden she "disfigures" herself independently of the particular content of the violated command (Taylor 1985, 91-92; 98; 103). This essential harm is what Taylor calls "the stain of guilt": "Guilt is a burden he has to carry, he cannot disown it, it must leave its mark upon him" (1985, 92; 103).

The harm that is non-essential to guilt experience is the particular harm associated with the content of the forbidden moral principle or command violated. Of course, on Taylor's account, the content of the principle or command can be related to the self or others. Taking her example, if I neglect the cultivation of my talents, then the non-essential harm in my guilt is the harm I do myself in not developing my talents. As we saw above, Taylor claims, then, that the self is harmed under *two* descriptions: (i) the essential harm that I am doing what is forbidden, and, in this case, (ii) the non-essential harms which follow from not cultivating my talents (e.g., physical, material or psychological harms). But, Taylor glosses over a relevant point when she claims that the harm I bring on myself in not developing my talents *constitutes* the wrong I have done myself. She then weakly asserts that, though there is this constitutive relation in

this particular case, “the specific way in which she thinks she has harmed herself is distinguishable from the harm that is the stain of guilt” (Taylor 1985, 103).

The constitutive relation is, I take it, that the particular way she thinks she has harmed herself is the instantiation of what she takes to be forbidden. That is, what Taylor calls the “non-essential harm” of guilt is the moral content or substance of the wrongdoing. Does it follow that, if there is no particular way in which she thinks she has harmed herself, there can be no essential harm, and without that, there can be no guilt experience? The invitation to this idea of a constitutive relation, even if she tries to restrict it to this particular example, points to a problem for Taylor. Without the particular way, i.e., the moral content or substance of the wrongdoing, in which a person thinks she has harmed herself or others, there may, equally, not be any guilt. Namely, it may be that non-essential harm is contingent on the grounds of its variable content but, it may be essential relative to the experience of guilt that there is such a non-essential harm *at all*. If the harm that is the stain of guilt is essential to it, and has neither a typically (if not necessary) constitutive relation with the specific way one thinks one has harmed oneself or another, then Taylor has to either explain (a) how most experiences of guilt do not require a non-essential harm, or, (b) give an alternative account of how, if not constitutive, the two harms are otherwise related.

Though Taylor does not offer resources to construct responses to these questions, it is relevant to note why I think she needs to make this distinction. Very basically, if guilt is essentially self-regarding as Taylor claims, and guilt is essentially connected with harm as she implicitly accepts, then she requires a form of harm that only the self can suffer in guilt. That is, she needs a form of harm that does not depend on the contingent self- or other-regarding content of a given principle or obligation. What results is the idea that the substance of the wrongdoing seems to have become only the essential harm; so the moral substance, the substance of the wrongdoing, what surely should be the focus of guilt or remorse, has gone missing. Taylor has entered an interesting area of how to understand harms but, as the next chapter will explore, Raimond Gaita’s

distinction between natural harms, and the significance of those harms, presents a better way of charting it.

2.3.4 Causal responsibility

The essential features of Taylor's self-regarding guilt so far outlined are the acceptance of authority, the essential thought ("what I am doing is forbidden"), and the essential harm (the person disfigures herself in doing what is forbidden). These features, taken together, make it clear not only how Taylor is influenced by the morality system, but also how her conception of guilt is "an emotion experienced in the face of an abstraction, the moral law, which has become part of the subject himself" (Williams 1993a, 219-220). The next essential feature is causal responsibility, which, in a move away from the morality system, Taylor takes to be both necessary *and* sufficient for guilt. In what follows, I first explore how Taylor arrives at the sufficiency of causal responsibility for guilt, followed by a discussion of how she argues for the necessity of causal responsibility.

The starting point for Taylor's discussion is the idea that guilt, unlike shame, has a privileged connection with responsibility: "It is indeed true to say that when feeling guilty but not when feeling shame I must think myself responsible for the relevant state of affairs" (1985, 90-91). It is, I take it, open to debate whether the claim regarding shame is correct. The claim I am interested in, however, concerns the alleged necessary connection between taking oneself to be responsible for the relevant state of affairs, and feeling guilt. To appropriately/justifiedly experience guilt a person must believe that *she* was responsible for bringing about the state of affairs. This is, I take it, expressive of the idea that beliefs reflecting the perspective of (moral) agency are necessary for any justified experience of guilt (Deigh 1988, 392).

Standardly in moral philosophy, what is taken to ground the necessary belief in one's responsibility as expressed in guilt is the requirement that one's action meet the following characteristic conditions of culpability (in retrospective moral assessment): whether what one did was done intentionally/voluntarily, carelessly or negligently, and/or through culpable

ignorance. Namely, there must be grounds for justified blame.³³ Once again, we encounter the idea that the conditions of blame, as second- and third-personal expressions of moral responsibility, are mirrored in guilt, as a first-personal expression of moral responsibility. But—and here is where Taylor makes her unusual but critical move—she claims that one need not meet the conditions of culpability to experience guilt; one can experience guilt for a terrible state of affairs one brought about purely *accidentally* so that blame is not warranted. Taylor writes,

I may but need not think of myself as having intentionally or negligently done or omitted to do what I feel guilty about; I may but need not think that I could and should have acted otherwise and so am blameworthy in the accepted sense. Normally, perhaps, I do think myself blameworthy in this sense when feeling guilty, but not necessarily so (1985, 91).

We can, in other words, feel guilt for causing bad things blamelessly. With this distinctive move, Taylor is separating the grounds for guilt from the grounds for blame (Greenspan 1992, 287).

To support her claim, Taylor presents a version of Williams' lorry driver: a driver knocks down and kills a child she could neither have seen nor avoided (1985, 91). The driver was not at fault; she was not negligent or careless, and she could not have reasonably foreseen that the child would leap out in front of her car. She would not be held legally responsible. Moreover, following the standard moral philosophical picture, there is no sense in which she could be justifiably blamed by others, and therefore cannot be held morally responsible for the death. And yet, the driver feels guilt. That is, she takes herself to be responsible for the death. Taylor wants to accommodate these extreme cases, and this leads her to claim that the belief in one's causal responsibility with respect to the relevant state of affairs is *sufficient* for intelligible feelings of guilt.

The immediate objection from an advocate of the standard picture would of course be to claim that, although Taylor is including such extreme cases of

³³ See Wallace (1994; 2013) for a respected Strawsonian example. With regards to his non-purified conception of the voluntary, Wallace is a not full member of the morality system. But, as Paul Russell (2013) argues, though there are merits to Wallace's account, he is still problematically operating with the morality system's narrow conception of moral responsibility.

bad moral luck in her conception of guilt, it does not follow that guilt is *rational* or *appropriate* in such cases. In light of this, the objection continues, she cannot be advocating a separation of, or asymmetry between, the grounds for rational blame set by the conditions of culpability, and rational guilt. Alternatively, one might appeal to R. Jay Wallace's distinction between causal and moral responsibility: "the actions of morally responsible people are thought to reflect specially on them as agents, opening them to a kind of moral appraisal that does more than record a causal connection between them and the consequences of their actions" (1994, 52). That is, in order for guilt to be a rational expression of moral responsibility, it must involve some kind of moral appraisal of the agent. Typically, moral appraisal, when directed at wrongdoing, is grounded in conditions of culpability.

Unfortunately Taylor has nothing to say on these issues. The only reply I can think to make on her behalf relates to her methodological qualification: she clearly states that she is not interested in normative standards of assessment for the emotions of self-assessment, but with what best explains them. In this way, Taylor is not interested in denying the rationality of guilt in such cases; she is simply taking the moral phenomenology of such cases to be instances of the experience of guilt. What, positively speaking, this permits is making clear the intelligibility of the lorry driver feeling guilty—it is a further question whether one takes it to be rational. Moreover, it leads one to ask the following question: without the assumption that in order for guilt to be "rational" it must meet the conditions of blame or culpability, can we understand the possibility of guilt (or remorse) being appropriate, and intelligible? As I will explain in Chapter 3, Raimond Gaita certainly thinks so. By the end of this thesis, I hope to show, at the very least, that this is a question moral philosophers ought to be asking.

What does Taylor mean by "causal responsibility"? Does she mean that though the driver could not have stopped the death, she was nevertheless instrumental in bringing it about as a hurricane can be said to be causally responsible for the flooding? Does the condition of causal responsibility require the agent to see a fairly direct causal connection between her deed and the terrible state of affairs, or can the causal connection be at a further remove? For example, one's actions conditioned the possibility for a terrible state of affairs to

arise without one directly causing it or participating in it. How far does causal explanation reasonably stretch in connection with questions of responsibility? Taylor does not offer much by way of what she means by “causal responsibility”. Her intention seems to be to use “causal responsibility” to suggest taking oneself to have been directly instrumental in bringing the situation about. But, what differentiates the case where an agent takes herself to be causally responsible without any feeling of guilt? On Taylor’s account, the difference will be explained by the absence or presence of the essential thought: that what she did, whether done voluntarily, negligently or through culpable ignorance, was forbidden.

Taylor takes her position on causal responsibility to be supported by the facts that “guilt cannot be vicarious”, and, “feelings of guilt cannot arise from the deeds or omissions of others” (1985, 91). She draws on what she takes to be a common contrast between shame and guilt: it is often thought that we can feel shame vicariously. Does Taylor mean vicarious by association, e.g., “I feel ashamed that my son stole from me”? Or, does she mean sheer contagion, e.g., “The very idea of *x* fills me with shame”? Given that her supporting example is of a mother’s relation to her son’s misdeeds, Taylor presumably means “vicarious by association”.³⁴ Though the mother may have failed her son as a parent, and so can intelligibly feel guilt over *this*, Taylor claims, the mother cannot intelligibly feel guilt over her son’s current misdeeds *per se*. She can, however, feel shame. “His own agency has broken whatever causal chain there may be between his defective upbringing and his present behaviour” (1985, 92). Though she may feel as though, in her capacity as a parent, she is responsible for her son and his actions, her son is sufficiently independent agentially to condition his status as being responsible for his own actions.

The point of this example, I take it, is to illuminate how the belief that one is causally responsible reflects the perspective of agency. The resistance in the possibility of feeling guilt by imagining the guilty feelings of another supports the idea that there is no sharing of one’s or anyone else’s guilt. Guilty feelings, in other words, are distinctively one’s own. Taylor’s appeal to this

³⁴ Enoch (2012) uses this example in his account of what he calls “penumbral agency”; what he calls “taking responsibility” in cases where one is not directly causally responsible for bringing a regrettable state of affairs about, but where one bears some ethically significant relation to it.

example reveals not just these common features of guilt but the idea that guilt has a necessary connection with cause. That is, without taking oneself to have played a causal role in bringing something about there can be no beliefs that reflect the perspective of (moral) agency, no question of responsibility and no experience of guilt. Together, what these facts about guilt and the example serve to reveal is that Taylor is also making the claim that causal responsibility is *necessary* for guilt.³⁵

Unfortunately, the following example is what Taylor offers in support of her argument: “If I feel guilty about my privileged position in society due to circumstances of my birth then I see myself as an agent causally involved: it is *my* birth which has brought about the state of affairs which is my privileged position” (1985, 91). The example does not function as Taylor intends it to. Either it undermines the claim that causal responsibility is necessary for feeling guilt, or Taylor needs to alter her project. Clearly, the man did not cause his birth, and its associated privilege. That the man was born into privilege and social advantage is a matter over which he had no agency, and there is no intelligible sense in which he can reasonably consider himself as causally responsible for his “in-built” advantage. On Taylor’s own terms, the man thinking himself an agent of injustice by virtue of being born into privileged is at best far-fetched.

We can make sense of this case on Taylor’s terms by re-describing it: that the man has acted in ways to use his advantage to his benefit at the expense or disadvantage of others less fortunate and, in so doing, has perpetuated the unjust structures of his society. Under the original description it makes better sense to say that the agent feels guilt at being a beneficiary of injustice (Deigh 1988, 392). I take it that this is a more plausible explanation of the case, and it brings out the fact that the case belongs to a different group of examples altogether: those in which one has feelings of guilt for things that are not one’s doing *at all*. In this group are such cases as the phenomena of survivor’s guilt and, in some instances, collective guilt. Survivor’s guilt is often taken as a clear case of irrational guilt owing to its complete lack of connection with causal

³⁵ See Kenny (1963, 193) and Williams (1993a, 56-57).

responsibility.³⁶ Given Taylor's stated investment in capturing guilt irrespective of characteristic rational and irrational borders, she needs to make sense of cases such as these; and yet, without the necessary causal responsibility, it looks as if Taylor's account simply rules out the intelligibility of such cases. If Taylor were to recognise that her case does not support her claim about causal responsibility as a condition on guilt, then she needs to consider the possibility that her claim, which she presents as a claim about intelligibility, really ought to be about guilt's justification.

2.3.5 Repayment

In analysing the final essential feature of guilt, "repayment", Taylor, again, takes something that is commonly thought to be essential to other-regarding guilt, and argues that it is not. Taylor takes aim at the idea that making reparations is essential to guilt. She agrees that, if the violated command was other-regarding, then making reparations may be one mode of response but accepting punishment, for example, is also a possibility (Taylor 1985, 90). Moreover, Taylor affirms, there can be cases where reparation is not even a possibility, or where there is nothing to repair (e.g., incest and parricide are not actions that Oedipus can make up for or undo). My formulation of her argument is as follows: though other-regarding reparation may feature in many cases of guilt, other-regarding reparation need not be present in all cases of guilt, therefore, other-regarding reparation is not essential to guilt. However, it is, as we shall see, essential to her account of remorse.

What is essential to every experience of guilt, according to Taylor, is "that the agent sees himself in a position where repayment is due" (1985, 103). Seeing oneself as in a position where repayment is owed is explained by one's taking oneself to have been "directly instrumental in bringing about the state of affairs" (Taylor 1985, 90). One's causal responsibility in guilt ought to be understood, according to Taylor, from a debit and credit point of view of punishment. That is, having brought about the misdeed or caused the state of affairs one thereby owes payment or credit for the imbalance caused by the

³⁶ See Velleman (2003, 246) for an attempt to argue, on separate grounds from standard guilt, for the rationality of survivor's guilt.

transgression (Taylor 1985, 93). The repayment can take the form of reparation of the damage, accepting punishment or retribution, or even being pardoned or forgiven. But, regardless of its form, Taylor's central contention is that it is essential to every experience of guilt that the agent recognise that, in doing what is forbidden, she has put herself in a position where she owes some form of payment to restore the imbalance brought about by what she did.

Further to this, Taylor explains that, "It is also because he thinks himself responsible in this sense [causally responsible] that we can account for the effect the deed or state of affairs has on him. The thought involved here is not so much: "I have done this terrible thing *to him*" but is rather "*I* have done this terrible thing to him"(Taylor 1985, 92). Taylor's emphasis on the importance to the agent that she is the one who did it is correct. For, owing to Taylor's dualistic conception of the distinction between self-regarding and other-regarding, and her resultant picture of guilt as essentially self-involved, she cannot recognise the possibility of guilt sometimes taking relational form e.g., "*I* have done this terrible thing *to him*". Once again we encounter Taylor's egoistic, self-involved picture of guilt that prevents the thought that guilt could be an emotion of self-assessment that is relational; that is, between an agent and a victim.

My critical interpretation is supported by Taylor's remarks regarding what motivates repayment. Given that the essential harm in guilt is the harm one causes oneself in doing what is forbidden, it not unsurprisingly follows, for Taylor, that an agent is motivated to actuate repayment by *her* disfigurement, and the burden of guilt *she* carries. In feeling guilt, Taylor claims, she sees herself as disfigured and feels uneasy in herself. The uneasiness supposedly explains the painfulness of guilt feelings (Taylor 1985, 98). In order to return to extirpate the painfulness, the person feeling guilt is motivated to remediate her disfigurement and uneasiness by lifting the burden and restoring her balance. Repayment is the guilt sufferer's means to the end of freeing herself from the burden of guilt. According to Taylor, the motivation to repayment in guilt is, in all cases, reducible to an egoistic motivation to restore one's sense of disfigurement and uneasiness. For Taylor, an agent's feeling of guilt is always

fundamentally motivated towards repayment (or reparation) by a desire to restore *herself*.

What this means for cases where an agent feeling guilt acts to repair the harm caused to others through her actions is that she is only incidentally “righting a moral wrong” (Taylor 1985, 97). If repayment is always motivated by self-interest then, clearly, Taylor is certifying guilt’s status as *moral* only insofar as it relates to the authoritative abstraction “what is forbidden”. Sensible to the objection that the harm done to the victim has all but gone missing from her account of guilt, Taylor attempts to counter-balance the egoism of her picture of guilt by contrasting it with its complementary counterpart: namely, remorse conceived as fundamentally other-regarding. It is to her account of remorse that I now turn.

2.4 Taylor’s account of remorse

In the introduction to this chapter, I explained that Taylor takes guilt and remorse to be distinct emotions warranting separate accounts. On my diagnosis, her basis for this claim is her view that guilt is an emotion of self-assessment whereas remorse is not. As we saw above, Taylor takes any other-regarding aspects of guilt to be either non-essential to it and/or reducible to egoistic self-involvement. What begins to emerge, most obviously in her picture of repayment, is that Taylor tends to equate the fact that guilt can be understood as an emotion of self-assessment with the idea that guilt is self-involved. What results is an extreme picture of guilt as an essentially self-involved emotion that is only moral insofar as it relates to “what is forbidden”. This picture is countered by Taylor’s account of other-regarding remorse, which shares many key features of her self-regarding conception of guilt, and takes on others that Taylor rejected as non-essential to guilt.

Following the morality system, Taylor draws out her account of remorse by contrasting it with regret. Specifically, she is concerned with arguing against a picture of remorse that reduces it to regret operating within a strictly moral domain (Taylor 1985, 98). One of the thoughts motivating such a picture is the widely accepted idea that remorse can be defined as a feeling of deep regret felt

over a sin or moral wrong (Taylor 1985, 98). Taylor agrees with the connection made between remorse and moral wrongs, but is interested in arguing for remorse as irreducibly distinct from regret. As I will show, in doing so, Taylor creates a picture of regret that amounts to a mere artefact of remorse. And, as I will show in the final section, Taylor also goes wrong in not recognising that her account of remorse partakes in key features of her account of guilt, in combination with those she argued were non-essential to it. To make clear how Taylor constructs remorse as distinct from regret, at the same time as indicating shared features with guilt, in what follows I critically reconstruct Taylor's lines of distinction between remorse and regret in the following terms: scope, causal responsibility, the view the agent takes of the action and its consequences, and connection with reparative action.

2.4.1 Scope

In the explication of the morality system's conception of remorse, the characterisation of its scope is accepted as being strictly delimited by "the moral". Taylor accepts the morality parameter on remorse, and contrasts it with the scope of regret, which is generically "felt about what is in some way undesirable" (1985, 98). The scope of intelligible regret, therefore, ranges across the broad category of whatever may be "undesirable" including (but of course not limited to), what is morally undesirable. Taylor gives two supporting examples: feeling regret over the passing of the summer, and Hamlet's regret "that circumstances had forced him into a position where he had to act against his own nature" (1985, 98). If we follow Taylor, and accept that one can intelligibly feel regret over the passing of the summer, then the object-range of regret extends to events or occurrences that do not involve human agency. Arguably, even if we accept such events as intelligible objects of regret they are surely exceptional. That is, the more prevalent, and interesting, domain of regret is one wherein human action is intelligibly involved; where, as mentioned in the previous discussion of causal responsibility, the inevitable questions and difficulties of attributing causation to human agents arise.

2.4.2 Causal responsibility revisited

As in guilt, Taylor claims that an agent must believe that she is *at least* causally responsible for a state of affairs in order to feel remorse for it. Causal responsibility, as in guilt, is necessary for remorse. Taylor writes, “Regret but not remorse can be felt about an event for which the agent does not take herself to be even just causally responsible” (1985, 98). Following the elaboration of causal responsibility in guilt, I will assume that Taylor is still taking it to mean “directly instrumental in having brought something about”. Taking regret first, if regret can be felt over events that do not involve human agency, then clearly regret does not require an agent to take herself to be causally responsible. But, if the regret is focused on human activity, it is reasonable to assume that an agent experiencing regret can take herself to have been directly instrumental in bringing something undesirable about. I may, for example, regret forgetting to leave a gratuity for the waiter after receiving excellent service at a restaurant. Perhaps, in an arguably ethically salient direction, I affirm that, next time I am at the restaurant, I’ll remember to leave a gratuity. What the example illustrates is the possibility of regret in instances where an agent takes herself to be causally responsible for bringing about the regretted event or situation without it being an instance of remorse. In this connection, the morality system can claim to have solid explanatory power: there was no broken moral obligation, and so no *moral* response is required, or even intelligible. Only some form of non-moral regret would be an appropriate response in such a situation. As we shall see, Taylor emulates this construction.

2.4.2.1 Taylor’s problematic Hamlet example

Immediately following the necessity claim regarding causal responsibility and remorse, Taylor presents the example of Hamlet. In what follows, I will assume that the example is supposed to offer support for the claim. She writes, “Hamlet regretted that circumstances had forced him into a position where he had to act against his own nature. He could not, seeing the situation in this way, have felt remorse” (Taylor 1985, 98). Let us interrogate Taylor’s claim that Hamlet’s lack of remorse for killing Polonius is explained by

his view that the murder and usurping of King Hamlet pushed him into acting in ways contrary to or outside his nature. On this interpretation, Hamlet does not see the event *as an action of his* on the grounds that he was not himself when he acted. Moreover, given that Hamlet was not involved in bringing the original circumstances about, he does not take himself to have precipitated the change in his nature.

Taylor's idea seems to be that neither at the level of direct action, nor at the level of conditioning circumstances, does Hamlet take himself to be even causally responsible for the actions performed in said state, and that this is why he can feel no remorse. But further analysis of the Hamlet example swiftly reveals this cannot be right. Clearly, there is a sense in which it is correct to say that Hamlet *caused* Polonius's death. Indeed Hamlet himself clearly registers some significant causal role in the death, for his offering an explanation is itself a recognition of a need to account for something one brought about, even if that amounts to a refusal of fuller responsibility (Williams 1993a, 57). For this reason we should not accept Taylor's claim that, it is the denial of causal responsibility that motivates Hamlet's denial of moral responsibility; rather it is the manifest fact that he was not in his right mind when he acted.³⁷ The refusal of responsibility happens, as we might put it, at the level of *state* and not *cause*; the reason for denying moral responsibility is that he was not fully sane, and therefore, not himself when he acted. This is why he does not see the action as an action of *his* in the manner necessary for remorse. I conclude that the example fails to support Taylor's claim that taking oneself to be causally responsible is necessary for intelligible remorse. All it shows is that one can acknowledge a significant causal role in bringing something about whilst still refusing responsibility for it at another level.

³⁷ The idea of "not being oneself or one's usual self" when one acted is invoked by Williams in his discussion of Agamemnon's response to his taking of Briseis from Achilles (1993a, 52-54). Though not his usual self, Agamemnon accepts responsibility for the action, and, in doing so, Williams writes, "He is not dissociating himself from his action; he is, so to speak, dissociating the action from himself" (1993a, 54).

2.4.2.2 Remorse, guilt and causal responsibility

A partial answer to my question can be found if we go back to some of Taylor's remarks on guilt and causal responsibility. Taylor agreed with the standard view that, in order for a person's guilt to be intelligible, she must believe that *she* was in some way responsible for bringing about the state of affairs. But, unconventionally, Taylor's interest in accommodating cases of unavoidable wrongdoing led her to claim that taking oneself to be causally responsible was sufficient for feeling guilt. In such cases, the belief in one's causal responsibility was assumed to sufficiently reflect the perspective of agency that is necessary for every experience of guilt. This, of course, does not preclude Taylor acknowledging that beliefs which reflect the perspective of agency could, and do, enter at other levels and in more complex ways, e.g., in cases where I should have known better than to act in the way that I did, which resulted in the terrible state of affairs.

Should we understand Taylor as doing the same regarding causal responsibility in remorse? After claiming that causal responsibility is necessary for remorse she writes, "Remorse is always felt about an event *which the agent sees as an action of hers*" [my emphasis] (Taylor 1985, 98). As I interpret this, Taylor is making the same claim regarding remorse as she did in connection with guilt: that taking oneself to be causally responsible reflects the perspective of agency. But, as suggested by the gratuity example, an agent can see that she was directly instrumental in bringing something undesirable about and feel regret, yet not remorse. Equally, as my analysis of the Hamlet example showed, he recognised a causal connection but dissociated himself from the action at the level of his state of mind. These considerations, therefore, raise doubts regarding the meaning of causal responsibility on Taylor's account. It is not, as Taylor previously claimed, equivalent to merely recognising that one's action was directly instrumental, for not even on her account is the recognition of a causal connection sufficient for an agent to see the action as hers.

Taylor is correct in thinking that being a cause is a necessary condition for the attribution of responsibility (whether to oneself or by others) (Williams 1993a, 56). Without the belief in one's causal role there can be no belief in one's

responsibility, and thereby no experience of remorse. But complexities arise when one can, for example, take oneself to have caused an event but one may not regard oneself as morally responsible for it, as in Hamlet's case. The agent seems to acknowledge a causal role in bringing some wrongdoing about, which may not be sufficient for *seeing the action to be one's own* in the way Taylor implies.

Given that causal responsibility is necessary for remorse as it is for guilt, is it also sufficient on Taylor's view? Does the intelligibility of remorse, like guilt, extend to instances of unavoidable wrongdoing? Unfortunately, nowhere does Taylor address these questions. The discussion of the next feature suggests that Taylor takes remorse to be appropriate only in cases where an agent has deliberated. However, she does not answer the question of how one might respond to non-defective deliberations that coincide with unavoidable wrongdoing.

2.4.3 The view an agent takes of her action and its consequences

Taylor elaborates the view an agent takes of her past action and its consequences in order to explain what she takes to be a notable difference between regret and remorse: "If she feels remorse then she wants to undo the action and its consequences which cause the remorse, but when feeling regret she need not think that she would undo the action if she could" (1985, 98-99). Here, I interpret Taylor to be making two claims: (i) it is not a necessary feature of regret that an agent even have the thought, let alone the desire, that she would undo the relevant action and its consequences if she could, and (ii) it is a necessary feature of the experience of remorse that an agent wants to undo the relevant action and its consequences, even if she cannot. But, as I will argue, this alleged contrast between regret and remorse is not plausible, for an agent may quite intelligibly feel remorse for things he has done even while not wanting to undo them—for instance, if they were a means to an over-ridingly valuable end. In pursuing Taylor's arguments in this connection, the artificiality of her conception of regret is perspicuously revealed.

The first claim allows for the possibility that, in regret, one might not even think, let alone have the desire, to undo the action and its consequences. To support the claim, Taylor discusses the compatibility of feeling regret over an action whilst accepting it as “all things considered the right thing to have done” under the same description.³⁸ Her example is of an employer who “regrets sacking the employee because the girl was so easily crushed, but she had to be sacked, nevertheless, because she was so inefficient” (Taylor 1985, 99). The employer sees the action as “necessary and beneficial”, and yet, she feels regret because the action had the disagreeable aspects of having to sack an employee, and the employee feeling crushed. In such cases, the agent does not see the disagreeable aspects as being outweighed by the good of the action. Namely, the good of the action relative to its consequences is viewed as overriding the disagreeable aspects. To view one’s action as “all things considered the right thing to have done”, Taylor implicitly assumes, would be incompatible with the desire to undo the action and its consequences.

Taylor refers to the compatibility of feeling regret and seeing one’s action as all things considered the right thing to have done as “acceptance”. Acceptance, I take it, refers to one’s willingness to tolerate the disagreeable aspects of an action owing to one’s seeing the action and/or its consequences as over-ridingly good. Again, Taylor assumes that to see the action and/or its consequences as all things considered the right thing to have done, and thereby accept associated disagreeable aspects, means that one would not have the desire to undo the action and/or its consequences. But, Taylor neglects the possibility that one can accept the disagreeable aspects as part of morally right action, and yet have the desire or wish to undo the disagreeable aspects. As we saw in the discussion of the morality system and tragic moral dilemmas, though Agamemnon takes himself to have done the morally right thing in freeing the fleet to sail, to suggest that he would not and/or should not have the desire or wish to undo his daughter’s murder is, as I shall discuss below, shallow (and absurd). This suggests that by acceptance Taylor means something closer to

³⁸ Taylor uses “necessary and beneficial” and “all things considered the right thing to have done” interchangeably.

“erase the moral importance or significance of the disagreeable aspects” (i.e., there is no moral residue).

Taylor does not recognise these issues, and goes on to conjecture that, “Perhaps regret always implies acceptance of what has been done. It had to be done although there were unfortunate or disagreeable aspects to the deed” (1985, 99). If we take seriously the idea that, regret entails acceptance, this pushes in the direction of claiming that in most cases of regret over one’s past actions, the agent views the action as all things considered the right thing to have done. If this is the case, then, on Taylor’s account, the desire to undo the action and its consequences will be characteristically incompatible with regret.

It seems obvious, however, that regret is often accompanied by the wish that one could undo the action and its consequences (admittedly, a possibility left open by (i)). I might drop and break my grandmother’s cherished antique vase. I feel regret that I broke it, and wish I could undo what I’ve done. But, Taylor would reply, breaking the vase was an accident, and not the result of a deliberation. Though Taylor does not go all the way in asserting that regret does entail acceptance, the strength of the suggestion reveals a tendency in her account to conspicuously ignore the compatibility of regret and the desire to undo one’s actions and/or its consequences. I suspect that this is owing to the force of the distinction Taylor wants to make between regret and remorse in relation to the necessary possession of the desire to undo one’s action and its consequences.

The stark contrast Taylor intends is evident when she writes, “Remorse...never implies acceptance. It is impossible to feel remorse and yet believe that overall it was right to act as one did”(1985, 99). Echoing the morality system on remorse, according to Taylor, if I view my action as all things considered the right thing to have done I cannot experience remorse. Of an agent experiencing remorse she writes, “The aspect of the action which causes remorse...is regarded by her as outweighing any possible good that may have come of it” (Taylor 1985, 99). So though some good consequences may, for example, have followed from their action, the remorseful agent *necessarily* sees any good as outweighed by the wrong or harm. This leads to the conclusion that the view the remorseful agent must take of her action and its consequences is

one whereby “all things considered it was *wrong* to act as she did”. With such a strong view of one’s action and its consequences, it is, therefore, unsurprising that Taylor takes the remorseful agent not simply to have the thought of undoing the action and its consequences, but the desire to undo them.

Taylor attempts to provide support for her picture in the example of Agamemnon (1985, 99). As I mentioned above, on Taylor’s reading, so long as Agamemnon sees the overriding good as the sailing of the fleet, and thereby views his sacrifice of Iphigenia as all things considered the right thing to have done, he cannot intelligibly experience remorse. If Agamemnon believes that he did the right thing sacrificing Iphigenia in order to set sail, it follows, according to Taylor (and the morality system), that he can intelligibly feel regret regarding the disagreeable or unfortunate aspect of his action: the murder of his daughter. To take Agamemnon as necessarily viewing the death of his daughter by his own hand as a disagreeable or unfortunate aspect of his otherwise right action sounds absurd and shallow. Surely, the expectation in such a case is that regardless of whether he views the action as all things considered the right thing to have done we, the audience (and his wife, Clytemnestra), expect him, at the very least, to have the wish that his daughter was not dead or the desire to undo the death of his daughter. This does not even enter into the question of how one responds first-personally to the fact that one has killed one’s own daughter. Taylor is sensible to this when she writes, “He could have felt regret, though unless this is qualified (deep, bitter) it would seem a rather inadequate reaction under the circumstances” (1985, 99). This begs the question of what “deep” or “bitter” regret might be in such a case if not an expression of remorse.

What makes the regret deep? Is it the non-moral quality of being merely disagreeable or undesirable? Would “deep regret” be some special form of moral regret that can be differentiated from remorse? A special form of moral regret that is neither remorse nor regret, and can accommodate just such cases of tragic dilemmas or perhaps even involuntary cases? As we shall see in Chapter 4, this is a common interpretation of “agent-regret”. However, in this context, if one goes in the direction of thinking Agamemnon’s deep regret is synonymous with remorse, it casts doubt on the following: (i) that remorse is necessarily incompatible with believing one’s action to be all things considered

the right thing to have done or that remorse necessarily involves believing one's action to have been all things considered the wrong thing to have done, and, concomitantly, (ii) the view that the desire to undo one's actions and its consequence is necessary for remorse.

2.4.3.1 The artificiality of Taylor's conception of regret

The declared strength of Taylor's position with respect to remorse, and its necessary connection with the desire to undo one's action, and its necessary disconnection with acceptance, can shed light on her less explicit comments regarding regret. When Taylor asserts that remorse necessarily involves the desire to undo one's actions and its consequences, she leaves the connection between regret and the desire to undo the action and its consequences conspicuously open. That this is significant is highlighted by Taylor's later conjecture that "perhaps regret always implies acceptance" which, though it may not strictly conflict with the thought that one would undo the action and its consequences if one could, it does conflict with the desire. This makes space for the implication that most cases of regret conflict with the desire to undo the action and its consequences. Of course, this implication does not fully exclude the possibility of regret and such a desire, but it suggests that it is the exceptional case. However, as the vase example aimed to show, regret and such a desire to undo the action are not simply an exceptional possibility, but are commonly taken to come together. That is, the desire to undo the action is arguably one common way in which regret is, and can be, expressed.

This leaves one to wonder what Taylor is trying to accommodate in her differentiation of remorse and regret. It seems that Taylor's project of contradistinction in combination with the strength of her claims regarding remorse push the contrast with regret to breaking point. It brings her close to implying a further claim: that the desire to undo an action and its consequences *necessarily* distinguishes remorse from regret. What crystallises is that Taylor's conception of regret is everything remorse is not: it is an artefact of her conception of remorse. This happens because she does not recognise that remorse is better understood as a *sub-category* or *species* of regret. Finally, even

though Taylor's language does register hesitancy, the distinction between regret and remorse she ultimately advocates is the one familiar from the morality system: if one feels only regret with respect to one's action and its consequences, one necessarily views the action as all things considered the right thing to have done, which conflicts with the desire to undo an action and its consequences. This is contrasted with remorse, whereby one necessarily views the action as all things considered the wrong thing to have done, and thereby has the desire to undo the action and its consequences. And, as we shall see in the next section, the desire to undo one's action and its consequences is what conditions remorse's characteristic response: reparation.

2.4.4 Reparation

Taylor has little to say on reparation. If regret entails the view that all things considered it was the right thing to have done, then, Taylor claims, "no action need follow from regret, or even need be expected to follow" (1985, 99). For Taylor, from the perspective of the agent viewing her action as right, she need not be motivated or expected to respond to any disagreeable aspects of her otherwise right action. Namely, at the level of the first-personal, there is only a contingent connection between regret and responsive action (whatever form it takes). But, she goes further when she adds that no action even need be expected to follow. Though it is unclear *who* exactly structures the possible expectation of a responsive action, there is the suggestion that there are grounds for thinking potential sufferers of the undesirable aspects of otherwise right action may not be (morally) justified in having an expectation of responsive action on the part of the agent.

This comes very close to the morality system's conception of moral obligation, and the impossibility of moral remainders or residues in right action. So long as one's action was all things considered the right thing to have done, there can be nothing morally relevant remaining to respond to. Any of the undesirable aspects of one's actions and its results in performing morally right action are regrettable, but they do not figure in the moral ledger. Though she may have caused bad things to happen through her right moral action, she is

thereby not morally required to respond with reparation for those non-moral bad things. Given Taylor's account of Agamemnon's sacrifice, one can assume that this extends to the extreme of tragic dilemmas. This reveals an assumption in Taylor's account (contra the essential feature of guilt, "repayment"): that reparation is a moral action necessary only in cases where one takes one's action to have been all things considered the wrong thing to have done.

If, in remorse, one necessarily views one's action as all things considered the wrong thing to have done, then, following Taylor, it has a necessary connection with reparation. Again, what Taylor takes to motivate reparative action in remorse is the desire to undo the action and its consequences. So, it is unsurprising when Taylor claims that remorse, unlike guilt, focuses not on the self but on one's actions and their consequences (1985, 98; 100). Moreover, she takes remorse to be unlike guilt in that it "is more outward-looking than guilt or shame, for, in remorse the agent is concerned with the effect of what he does on others" (Taylor 1985, 101; 105; 99-100). In this way, Taylor's account of remorse is as a primarily other-regarding emotion in response to wrongdoing, and, unlike guilt, not merely other-regarding on the contingent basis of the content of a moral principle or norm violated. This leads Taylor to claim that the other-regarding character of remorse is what distinguishes it from guilt as a "more genuine" moral emotion (1985, 101). In other words, that remorse preserves the moral substance of wrongdoing leads Taylor to consider it a more genuine moral emotion than guilt. It is in her account of remorse that Taylor locates concern for others affected by one's actions.

On the expectations of others with respect to reparation Taylor writes,

We do expect some sort of action from her who feels remorse, though of course we may expect in vain. She wants to undo what she has done, and although it is evidently impossible to do just that, she would normally be expected to try and do something towards repairing the damage she takes herself to have brought about. If she takes no such steps the claim that she feels remorse would be suspect (1985, 99).

The desire to undo one's action and the condition of responsibility are both important here. On the one hand, there is the idea that the desire to undo one's action and its consequences is necessary for remorse and yet, for a myriad of

reasons, the presence of this desire does not always result in reparative action, let alone effective reparative action. This is where the action and its consequences are such that there is no adequate form the reparation can take, i.e., no restitution that can adequately address the loss, or no sense in which the action and its consequences can reasonably be undone. But, inadequate action and/or inaction carries with it the threat of raising doubts about the authenticity of the agent's remorse. Namely, at the very least, some attempt at reparative action is also taken as a signal, and sometimes measure, of the genuineness of one's remorse. This is an important, and neglected insight, which, in the next chapter, I will take up in my sketch of an account of reparation. For the present, I turn to the final section where I take up Taylor's discussion of the relation between her conceptions of guilt and remorse.

2.5 Guilt and Remorse

On Taylor's account guilt and remorse are both first-personal expressions of responsibility for wrongdoing. Each emotional response requires the agent to take herself to be causally responsible for a harm or wrong and, thereby, in a position where repayment or reparation is taken as necessary. On the basis of Taylor's account we can say that guilt is differentiated from remorse in being an emotion of self-assessment, and essentially self-regarding in all its aspects.³⁹ However, in spite of this supposed line of distinction, given the high degree of similarity between her accounts of guilt and remorse, we can ask whether guilt and remorse really are as distinct as Taylor makes them out to be. That is, perhaps remorse and guilt are just the same emotion, or remorse and guilt are part of the same psychological field such that we can understand remorse as a form of guilt feeling particularly inflected by harm done to others. In what follows, I consider Taylor's claim that guilt and remorse can be experienced about the same event, which, once analysed, has the effect of

³⁹ Taylor considers Scheler's (1972) contrast in which guilt is a destructive emotion, and remorse is a constructive emotion. Though Taylor recognises that guilt is highly susceptible to self-preoccupation and indulgence, owing to her commitment to guilt as an emotion of self-assessment, she cannot coherently take it to be entirely destructive (1985, 101). And, though Taylor agrees with Scheler that remorse is a constructive emotion in its other-regardingness, she disagrees that it is redemptive or entails salvation owing to her separation of remorse from self-assessment (1985, 104).

reinforcing my scepticism regarding her assumptions of any significant differences between guilt and remorse.

On the basis of her account Taylor claims that “Guilt and remorse may be experienced about the same event” (1985, 100). Her example is of Macbeth and his supposed feelings of guilt *and* remorse in response to murdering Duncan: “Macbeth initially feels remorse as well as guilt, he wishes the deed could be undone...This is different from his guilt reaction; his inability to say ‘amen’ indicates that he is now marked as an outsider who has broken his relationship with God” (Taylor 1985, 100). Taylor’s use and interpretation of this example simply reflects her account. In Taylor’s interpretation of Macbeth’s response, we encounter her picture of guilt as an internal experience in the face of an absolutely authoritative abstraction, and as an emotion of self-assessment (e.g., Macbeth’s imagined change in status with respect to God). Under this kind of description, as expected on Taylor’s account, we do not discover any concern for the wrongdoing as it relates to Duncan. Presumably, according to Taylor, any concern for or with Duncan finds expression in Macbeth’s wish that the murder could be undone. The difficulty, for Taylor, is that in claiming that the first-personal expressions of moral responsibility, namely guilt and remorse, can be experienced about the same event, does not illuminate how, in light of their evident similarity, we ought to understand what makes them truly distinct emotions. If anything, claiming that they can be experienced about the same event puts greater critical pressure on this question.

Sensible to this line of objection, Taylor claims that remorse can be experienced without guilt, and guilt can be experienced without remorse. That guilt can be experienced without remorse is evident, for Taylor, from the examples where one feels guilt for not engaging in forms of self-improvement. It is less clear, however, what it can mean on Taylor’s account to experience remorse without guilt. On the character of remorse without guilt, she writes, “The wrong done need not present itself to her who feels remorse as forbidden, she need not think of herself as having disobeyed a categorical demand” (Taylor 1985, 100). It seems that what differentiates remorse without guilt is that the agent’s understanding of her wrongdoing need not find expression in terms of a transgression of the forbidden. Although this could merely be an instance of a

difference in levels of description, this is not specifically the case for Taylor. What this suggests, I take it, is that Taylor is assuming remorse can involve a different way of thinking about wrongdoing (1985, 100). On the basis of the idea that in remorse one need not think of one's wrongdoing in terms of the forbidden, Taylor claims that the beliefs involved in guilt and remorse are different and, therefore, that there is not a necessary connection between remorse and guilt. However, on Taylor's account, this seems to mean that there is no relation between these two ways of thinking about wrongdoing.

In part, I take the idea of there being two different ways of thinking about wrongdoing to reflect Taylor's commitment to her two distinct criteria of what makes something *moral*: (i) its relation to abstract principles or norms one takes to be authoritative, and (ii) its other-regardingness. In connection with first-personal wrongdoing we get the following formulation: there is moral wrongdoing in terms of the violation of abstract principles and norms viewed as authoritative, and there is moral wrongdoing that is concerned with how one's actions and consequences relate to harm done to others. The way it comes out on Taylor's account is that remorse is a first-personal expression of moral responsibility that focuses on harm done to others, and guilt is a first-personal expression of moral responsibility that focuses on a broken moral principle or obligation and the harm this causes oneself. Conveniently, this tracks Taylor's distinction between essential and non-essential harm.

Taylor, I take it, assumes that one's understanding of one's wrongdoing in remorse emphasises other-regarding "non-essential harm" (the concrete harm), whereas guilt emphasises what she calls "essential harm" (the harm one suffers in breaking a moral principle). Specifically, I take Taylor to have assumed that the distinctive other-regardingness of remorse means that the agent's understanding of her wrongdoing involves a more direct focus on the concrete harm suffered by others (i.e., on the non-essential harm). But, as previously discussed, the concrete harm is surely constitutive of the essential harm such that, any significant emphasis on one over and above the other is ultimately misleading or distorting. I think a better way to understand the agent's view of their wrongdoing is to understand the difference as one of emphasis; that is, whether the agent emphasises the breaking of the principle or

rule, or whether they emphasise the actual harm that constitutes the breaking of the principle. Formulating the difference as one of emphasis (rather than a clear distinction) draws out the point that it surely matters, morally speaking, that in cases where an agent harms others but is solely focused on the idea that one violated a moral principle, we might want to say that she is avoiding confronting the wrongdoing. Moreover, as we shall see in the next chapter, this formulation makes it possible to think of Taylor's guilt as a *corrupt* form of guilt, and not as representative of genuine or proper guilt.

This final argument supports the doubt that opened this section, that is, whether Taylor really can support guilt and remorse as being as distinct as she claims. Once Taylor claims that these first-personal expressions of moral responsibility for wrongdoing can be experienced about the same event, what the distinction seems to come to is a difference in the agent's emphasis on principles as opposed to people. At this point, there is nothing in Taylor that prevents the natural thought that remorse and guilt just are the same thing, or that remorse is best understood as a form of guilt feeling particularly inflected by harm done to others. Admittedly, these two options may eventually collapse into one another if, like Bernard Williams, one upholds the idea that,

A better candidate for what might be called "irrational" guilt is guilt felt simply at breaking a rule or a resolution, where there is no question of wrong to others or reparation. Robbed of those implications, guilt narrows down suspectly to a desire for punishment. It might then be helpfully replaced by what it should have been in the first place, shame (Williams 1993a, 93).

Regardless of whether one wants to follow Williams (and Gaita), it seems clear that, at the very least, Taylor cannot compellingly motivate her claim that guilt and remorse are properly distinct emotions.

2.6 Conclusion

Early in her discussion of guilt Taylor writes that, "in feeling guilt, I should think of *myself* as having harmed another applies, perhaps, to the most typical cases" [my emphasis] (1985, 87). As we saw in Taylor's arguments for

self-regarding guilt, she takes any concrete other-regarding harm in connection with guilt to be non-essential to it. But, it is not simply that other-regarding harm is non-essential to guilt. What Taylor claims leads to the conclusion that, in all cases of guilt, others and the harm caused to them, are relevant only insofar as they relate to the “forbidden” principle or norm violated, and to the possibility of the agent restoring her uneasiness or disfigurement in repayment or reparation. This is an example of what Williams identifies in his “disconnection with victims” objection: that “The victims may reenter, of course, in an account of what the subject has done in violation of that law, but that gives them, or thoughts of them, no more intimate connection to guilt than they have to shame”(1993a, 223). On Taylor’s account, any thoughts regarding harm caused to others, and any concern for reparation, are explanatorily reducible to the agent’s relation to, and violation of, the relevant principle, the essential harm this causes the agent, and her desire for self-reparation.

In her argument for self-regarding guilt, Taylor argued against a picture of guilt that takes it as essential to guilt that it is other-regarding. What Taylor does not recognise is that the other-regarding guilt she argues against ends up as her own account of remorse. First, there is the fact that Taylor conceptualises both guilt (whether self or other-regarding) and remorse as first-personal expressions of moral responsibility. But, more specifically, in Taylor’s formulation of the other-regarding conception of guilt, we find the agent seeing the action as hers, the agent’s attention focused on the victim, her thought focused on the harm caused to others by her actions or omissions, and the desire to make reparations for the harm. In her account of remorse, we find the agent seeing the action as hers, her attention on her actions and their consequences, which Taylor takes to mean a focus on those harmed by one’s action. And, of course, she considers the desire to make reparations essential to remorse. Is Taylor’s account of remorse not just the account of other-regarding guilt she rejects in favour of self-regarding guilt? I think it is, and I take it that the problem lies in the convergence of Taylor’s commitment to the morality system’s conception of moral authority, her interpretation of what it means for guilt to be an emotion of self-assessment, and her extreme dualism concerning the possibility of perspective being either self or other-regarding.

Taylor is problematically committed to the morality system's conception of moral authority, to the idea that what it means for an emotion to be an emotion of self-assessment is that its primary object is the self, and to an extreme dualism of perspective (either wholly self or other regarding). The convergence of these commitments results in an account that is both blind to the evident possibility that guilt and remorse are not distinct, and an account that cannot recognise the possibility of a non-dualistic, relational, perspective in remorse or guilt. Moreover, once Taylor's self-regarding guilt is seen to relate to other-regarding remorse, it leads one to question the moral adequacy of an account of guilt that focuses exclusive attention on the breaking of moral principles at the potential expense of the moral content, i.e., of the harm done to others. This is not to suggest that all accounts of guilt that focus on the breaking of moral principles should be understood as reducible to sheer self-concern, as mediating or disowning the moral content of what one has done. But, it can certainly put explanatory pressure on those accounts to explain how concern for the breaking of moral principles relates to or leads to the kind of reparation that reflects a genuine concern for the victim in guilt or remorse. An upshot of the contrastive picture is that one can see something I will consider in the next chapter where I take up Raimond Gaita's novel account of remorse: that a conception of guilt like Taylor's may be best understood as a *corrupted* form of genuinely other-regarding guilt or remorse.

CHAPTER 3

GAITA ON REMORSE

3.1 Introduction

There are few places in contemporary moral philosophy to look for a philosophical account of remorse, let alone one that is not the product of the morality system. Raimond Gaita offers a distinctive account of remorse that is, unusually, a product of a Wittgensteinian approach to moral philosophy, and not a product of the morality system.⁴⁰ Unlike Gabriele Taylor's account, Gaita's does not split guilt and remorse apart into self-regarding and other-regarding emotions. On his account, they are one and the same: "If we understand guilt-feeling to be a pained acknowledgement of the wrong one has done, then there is no significant difference between guilt-feeling and remorse" (2000, 34).⁴¹

Gaita argues that remorse, as a suffering recognition of one's guilt, has both self-regarding *and* other-regarding dimensions. Moreover, on his view genuine remorse is characterised by two somewhat opposing energies: it is at once essentially outward-looking in its concern with those harmed or wronged, and yet it is also depicted as radically inward-looking such that Gaita describes it as "radically singular". As I shall explain, a view that affirms such opposing energies without synthesising them risks pressing remorse, once again, into an artificially dualistic mould—though this time the dualism is internal to remorse rather than generated by way of an alleged contrast with guilt, as it was with Taylor. To avoid the internal schism created by the opposing directions of attention that Gaita rightly sees in remorse, he needs to move beyond an account that merely affirms one, then the other energy (now outward, now

⁴⁰ D.Z.Philipps, a Wittgensteinian, and H.S.Price (1967), argue for a short sketch of remorse that bears remarkable similarity to Gaita's eventual account. Gaita's work has also inspired other philosophers, e.g., in connection with his conception of remorse, see Steven Tudor (2001), and Christopher Cordner (2007; 2008). And, for a discussion of Gaita's conception of remorse in the philosophy of law see Anthony Duff (2011, 68-81).

⁴¹ It is worth noting that the assimilation is not entire. Gaita writes, "Another way of characterising remorse is to say that it is the recognition of what it means to be *guilty* of having wronged someone. That being so, there is not much difference between remorse and guilt feeling. There are differences—one hesitates to speak of a remorse trip, partly for the same reason that it comes less naturally to speak of neurotic remorse rather than neurotic guilt—but the differences are not so great" (2000, 4).

inward), and achieve instead a model that can synthesise them. I shall argue that he fails to do this, and that what is needed here is a model of remorse that depicts it as a *relational* emotion. Remorse, as I shall propose, does exhibit at once an outward and inward trajectory, and this is because remorse relates the subject and object of a wrongful action. This notion of remorse as a relational emotion incorporates both inward and outward energies at work in remorse.

This chapter is comprised of seven sections. The first section briefly introduces general features of Gaita's overall critique of moral theory, and his method—what sets him apart from the morality system or what he calls “the moralistic conception of morality”. The second section presents and discusses Gaita's account of remorse. The third section explores various corrupt forms of remorse, with the fourth section examining the special relation Gaita claims holds between remorse and moral understanding. The fifth section sets out Gaita's position against a narrow conception of moral responsibility that ties it too tightly with conditions of culpability, and concludes with Gaita's alternatively broad conception of moral responsibility. In the sixth section, I offer an account of reparation that takes advantage of some of the key materials made available in the previous sections. The seventh section presents a brief analysis of Gaita's most radical claim: that remorse ought to be at the centre of moral philosophy.

3.2 Gaita on moral theory and some remarks on method

When we ask what makes a principle a moral principle...an obligation a moral obligation—then I think we should seek at least some part of the answer in the kind of elaboration we give when we express most seriously our sense of what it means to wrong someone. Nowhere is that sense more sober than in lucid remorse. ...But now, if one puts in the mouth of the remorseful person many of the philosophical accounts of what makes an obligation a moral obligation or a principle a moral principle, of the nature of morality and of its authority, we get a parody. ... ‘My God what have I done? I have violated the social compact, agreed behind a veil of ignorance.’ ‘My God what have I done? I have ruined my best chances of flourishing.’ ‘My God what have I done? I have violated rational nature in another.’ ...Even if one thinks the parodies to be to some degree unjust, they point unmistakably to the fact that the individual who has been wronged and who haunts the wrongdoer in his remorse has disappeared from sight (2004, xxi).

This lengthy passage offers a rich and complex challenge to moral theory (as well as a rare example of the effective use of humour in moral philosophy). Typically, moral theory variously aims to codify moral thought, to reveal morality's *most* fundamental principles, organisation or structure, to orient moral value variously as conducive to human flourishing, or in relation to respect for our rational agency, or the maximisation of aggregate happiness, etc. This practice in moral philosophy, according to Gaita, tends to encourage a perspective placing moral obligations, rules, principles and conditions for the ascription of culpability “at the centre of its field of moral vision” (2004, xvii).

This moral-theoretical conception of morality, with its focus on underlying impersonal rules and principles, tends to lose sight of the original subject matter of moral thinking—as we saw in the previous chapter, our morally contentful thoughts and reactions. When Gaita puts the terms of accepted moral theories into the mouth of the remorseful person (as illustrated in the quote above), he effectively draws attention to the distorting dynamics of moral theorising as they relate to the subject matter of moral philosophy. As soon as the business of moral theorising is underway, Gaita suggests, there is the expectation and expression of a complete disconnection from the *imminent* perspective on that very subject matter; as though there are two absolutely distinct selves, one as an individual human being, and of the other as a philosopher (2004, 23).

The re-descriptions of the form and content of our moral thoughts and reactions in terms of the theoretical apparatus of a given moral theory, typically leads philosophers to distort, ignore and lose much, if not all, of the original concrete content.⁴² In the passage, Gaita sites one particularly significant loss in connection with remorse as a response to wrongdoing: “the individual who has been wronged and who haunts the wrongdoer in his remorse disappears from sight”. What this comes to is that, for Gaita, this narrow conception of morality encourages a picture of wrongdoing as constituted by the transgression of abstract principle—a picture familiar to us as a defining feature of the morality

⁴² See Chappell (2014) for a wonderful, very current, text that takes direct aim at the idea that philosophical moral theories, in their preoccupation with generalisations, typically erase or obscure the particulars, and their importance.

system, and which we detected in Taylor's account. And so, the concrete *harm done* goes missing.

This loss of concrete moral content in connection with remorse and guilt is familiar from Chapter 1, and my closing formulation of MS-remorse. There, I argued that MS-remorse suffered from a self-involved internalism, and attention directed to the moral law that could not capture a concern for those directly affected by one's actions. We also encountered a version of this loss of concrete moral content in Chapter 2, where I introduced Williams' version of this objection to the effect that guilt, when it is excessively purified and becomes a disposition to respect "the moral law", loses what makes it morally important: its special connection with victims (1993a, 223).⁴³

Among other things, Gaita diagnoses these tendencies in moral philosophy as reflecting a wish to "rid oneself of all that makes one vulnerable to it [sentiment] and yet retain the content of one's thought...To wish to be free of all that makes us vulnerable to cliché, banality, sentimentality" (2004, xxxiv). For Gaita this is not a question of romanticising emotion at the expense of reason. Rather, it is about the impossibility of fully escaping one's individual moral point of view in doing moral philosophy, and, the inseparability of thought and feeling as it relates to deepening moral understanding—the task of moral philosophy. Accepting these facts, for Gaita, is not a matter of straightforwardly losing a kind of independence or impersonality one seeks in critical philosophical reflection. As part of his response, Gaita advocates, borrowing from Wittgenstein, the idea of critical grammars of moral response. The critical grammars of moral response helps us to judge which responses are lucid, honest or genuine as opposed to those that are cowardly, sentimental, self-indulgent, banal, or false semblances. The various critical concepts, constitutive of any given "critical grammar", help us to recognise "what it is to think well and badly in this or that realm of inquiry of reflection" (2004, xxxv).

⁴³ Is the narrow conception of morality not just the morality system? Certainly, as will become clear, Gaita is arguing against a narrow conception of morality or "the moralistic conception of morality" that bears much similarity with Williams' morality system. Nevertheless, Gaita eschews this label partly because the narrow conception of morality, in his terms, is a corruption of morality, and, partly because he thinks it would be extremely difficult, and artificial, to extinguish our use of "morality" (2004, 244). The later Williams, I think, would be sympathetic to that claim, see "Moral luck: a postscript" (1995e, 243).

Moreover, as we shall see later in this chapter, according to Gaita, the objects of our moral response, and the response itself, have a constitutive relation, such that, there is no independent moral fact that, once grasped, orients our judgement of others' responses. The genuineness or authenticity of another's response can teach us about the significance of what that individual is responding to (Fricker 2009, 317).

From this perspective, Gaita takes the starting point in moral philosophy to be the positive sympathetic bonds between human beings. Though he shares this starting point with Williams (by way of Hume), Gaita quickly turns his focus to the breaking of these positive bonds. In so doing, he is concerned to capture, at a more immediate level, the consciousness of the guilty or remorseful person (i.e., the moral phenomenology of remorse). In this way, giving philosophical weight to the "what-it's-like-ness" of moral experience is central to Gaita's philosophical descriptions and elaborations.⁴⁴ Rather unusually, he sources the primary descriptions not just from the typical sites of literature, and imaginary cases, but from testimony (2004, xii). Specifically, the testimony of others, as well as his own testimony of three personal, and formative, encounters. "The subject-matter for ethical reflection is primarily action and speech which has a certain authority, and when it is speech, it is by those of whom we say that 'they have something to say' because they speak with an authority that derives from the way they have lived their lives" (2004, 4).⁴⁵ Aside from asking the reader to trust the authenticity and authority of the testimony, Gaita himself engages in a constant process of critical reflection that questions the concepts and assumptions, which feature in his own descriptions and discussions, as well as, his interpretations of the descriptions of others. Having outlined Gaita's overall approach, I now turn to an analysis of his account of remorse.

⁴⁴ I am borrowing the phrase "what-it's-like-ness" of moral experience from Chappell (2014).

⁴⁵ See Gaita (2004, 140) for further elaboration.

3.3 Gaita's account of remorse

3.3.1 *The outward trajectory of remorse*

Gaita's account of remorse begins with the example of a Dutch woman's response to her role in the death of three Jews during World War Two. The woman had secretly housed the Jews in her home but, once she became involved in a plot to assassinate Hitler, she judged that housing the Jews presented too great a risk to the plot. Soon after leaving her home the three Jews were captured and sent to a concentration camp where they were murdered. Sometime after the war, in a television interview, the woman said that she hated Hitler for much, but above all for making her a murderess. This woman is, of course, no murderer; she did not kill the three Jewish people, nor order someone else to kill them, nor would she be held legally responsible for their deaths. Most others would not hold her responsible for their deaths, or remotely blame her for what happened.⁴⁶ Assuming we take her literally, must we regard her as making some kind of moral mistake?

According to Gaita, the best way to understand her judgement that she had become a murderer is not to characterise it as "irrational" but to understand the judgement as a reflection of her pained recognition of the *seriousness* of what she had done. This pained recognition is the suffering perception of her role in the harm done to those three Jews, and what she had become in light of it. *This*, Gaita claims, is a lucid expression of remorse. Remorse, Gaita writes, "is an awakening to the terribleness of what was done" (2004, 52). The awakening to the terribleness of what one has done directs one's attention *at once* outward towards the person wronged, and inward to focus on what, as wrongdoer, one has become. According to Gaita, the indivisible attention on the pain one has caused another in remorse is conditioned by the recognition of the reality of that particular individual.

⁴⁶ There will likely be others who do blame her because she can be construed as not meeting the epistemic condition on responsibility: she *could have known* that such a thing would happen.

The victim enters the remorse of the one who wronged her (irreducibly) as one who was wronged. If she was betrayed, for example, then that is an irreducible dimension of the evil she suffers, whether or not she suffers any of the natural harms which are normally consequent upon betrayal. If the one who betrayed her is to see that her victim suffered the evil of betrayal, she must find it intelligible that her victim is the kind of being who can suffer that kind of evil (2004, 77).

As I interpret it, there are two key points to extract from this passage: (i) it is necessary for genuine remorse that one finds it intelligible that “the victim is the kind of being who can suffer that kind of evil”, and, (ii) a distinction between “natural harms” and their “significances”.

The first point captures the condition that, in order for the wrongdoer to see the betrayal in the light of remorse, she must find it intelligible that the victim is the sort of being who can suffer that kind of evil or wrong. According to Gaita, the wrongdoer must recognise the reality of the victim as an *irreplaceable* human being. Gaita illuminates the condition with the case of an American slave owner. Though there is a sense in which the slave owner saw his slaves as physical human beings, Gaita claims there was an important sense in which he did not see them as “irreplaceable individuals” (2004, 60). So long as slaves were not human beings in *this* sense, it was unintelligible to the slave owner that slaves could suffer the kinds of evil white people could.

This point is further elucidated by Gaita’s introduction of the example of a slave owner who rapes one of the workers. In such a case, Gaita explains, unless he came to recognise the slave’s irreplaceable individuality, there would be no possibility of remorse. In contrast, the slave owner may come to feel remorse for raping a white woman, because he sees her as an irreplaceable individual. A similar point could be made in relation to any example where, though an act is considered morally wrong in a given culture, it does not intelligibly apply as a wrong to certain groups or people. Inasmuch as, a particular group, in a given culture at a particular time, does not have full moral status of a human being, then to that extent they cannot be perceived as being harmed or wronged by actions which, when done to others, would count as harm or wrongdoing. This, of course, does not exclude the possibility that others in that given culture did or do perceive members of those groups as being

human in the relevant sense. As I understand it, the central point of this condition is that it is necessary for the recognition of harm done, which grounds remorse. If one does not recognise a victim *as* someone who can suffer evil, then the suffering perception of remorse, which awakens one to the reality of another, is not possible. In construing this condition in terms of understanding and intelligibility, it follows, then, that if you are not seeing someone as irreplaceable, as in the case of the slave owner, it is not just a practical moral failing, it is also, inextricably, a moral *cognitive* failing.

The second key point I want to take from the passage is that the harm the victim suffers is not reducible to the natural harm. By “natural harm” Gaita means such things as bodily harm or the loss of one’s belongings such as money or property. The disvalue of such losses can be distinguished from the significance of the victim having been betrayed. The very useful distinction Gaita is making is between natural harms and the moral significance of those harms. That an action and its consequences were unjust, murderous, a betrayal, or disloyal, are integral to the harm done, and the harm suffered by the victim. This is an important distinction because, without it, the wrong of, say, a violent assault in which one’s legs are broken looks the same as the harm of slipping and breaking one’s leg by accident.

Given how important and useful I take this distinction between natural harms and their significances to be, it is worth further elaboration. Gaita writes,

If we are treated justly then we receive not merely certain natural benefits or goods, but also just treatment as a distinct and irreducible object of gratitude. And a person who is the victim of injustice suffers not merely a determinate form of natural harm, such as the wounds inflicted by torture, but also the injustice of it, which is a separate and irreducible cause of his torment (2004, 77).

As the quote makes clear, the distinction works in both positive and negative directions. Focusing on the negative, the victim of torture suffers from the physical assault but also from the injustice, and possible betrayal, of having to experience torture. The determinate form of natural harm caused by the torture does not alone explain or account for the harm done to the victim, and what they suffer in light of it. For the victim, Gaita argues, the injustice suffered

cannot be reduced to the physical violence; the evil suffered irreducibly involves the injustice of the harm.

To paraphrase the positive inflection, the injustice of it is a distinct and irreducible object of suffering. Clearly, Gaita takes the claim further, arguing that the suffering caused by injustice, murder, betrayal, theft are “separate and irreducible causes” of one’s anguish. Making this distinction allows us to make sense of cases where, for example, a person’s home is broken into and robbed. The goods are covered by insurance, and easily replaced. What the replacement of the goods cannot address is the victim’s sense of violation, fear and loss of a sense of safety in their home.

It is important to note that this distinction is not simply relevant with respect to the victim and what they suffer. The irreducibility of the significance of the harm done also affects the wrongdoer insofar as she has become someone who has done *that thing*, i.e., someone who has betrayed, murdered, tortured, or assaulted another. This helps to make sense of what Taylor failed to capture in her distinction between essential and non-essential harms: “the specific way in which she thinks she has harmed herself [non-essential harm] is distinguishable from the harm that is the stain of guilt [essential harm]” (Taylor 1985, 103). Arguably, what Taylor was attempting to capture was an idea I think we are now in a better position to understand: the idea that the moral significance of what one has done is not reducible to the natural harm caused. The wrongdoer’s “stain of guilt” is better understood in Gaita’s terms as the recognition of the full significance of what it can mean to become someone who has murdered, betrayed, stolen, deceived for personal gain. These ideas will take on further relevance when I take up Gaita’s characterisation of the “internal pole” of remorse, what he calls its “radical singularity”. With regards to the central distinction between natural harms and their significances, it will acquire further importance in the later sections on remorse as a form of moral understanding, and my sketch of a Gaita-inspired conception of reparation.

3.3.2 Remorse as not mere feeling

Remorse, as Gaita variously explains, is very demanding:

Remorse requires a sober collectedness, or perhaps, more accurately, it aspires to it as a perfection belonging to its nature. A person who is unhinged cannot be responsive to the demands of remorse, and someone who cannot be responsive to its demands cannot be in remorse, for to be in remorse is to be in disciplined obedience to its requirements (2004, 59).

Taken together, the idea that remorse has demands internal to it, and, that those demands require a disciplined obedience in response to them, indicate Gaita's conceptualisation of remorse as not a "mere feeling" or "mere attitude".

Remorse, for Gaita, is, amongst other things, a "disciplined remembrance of the moral significance of what we did" (2004, 59). Neither a fleeting feeling, nor an attitude one occasionally inhabits, remorse is a persistent thoughtful perception of the wrong one has done, and the suffering caused to the person wronged. A useful comparison here might be mourning. Grief or mourning is, characteristically, neither a fleeting feeling nor an attitude one occasionally inhabits. Rather, one is *in* mourning. Like remorse, mourning might be thought of, particularly in its most serious forms, as a demanding, persistent attention to the loss one has suffered, and its significance.⁴⁷ Similarly, remorse, on Gaita's view, is a sustained moral perceptual state.⁴⁸

Understanding remorse as a sustained moral perceptual state need not imply that it is a state one is perpetually in. When one feels remorse, it is as a sustained attention to the full significance of the harm done, which may recur, or be protracted over longer or shorter periods of time. Though Gaita does not refer to remorse explicitly in terms of time, there is the suggestion in his formulation that even once it has passed, remorse is a state to which one is perennially vulnerable to returning. This is supported by Gaita's claim that remorse is unlike other forms of serious suffering in that, if severe enough, it "sticks with us" (2004, 47). The idea that, when suffering is severe enough it

⁴⁷ Gaita does make connections between remorse, grief and love as forms of recognising the reality of an individual (2004, 47-48; 51-53).

⁴⁸ The combination of perceptual and cognitive elements makes Gaita's account an example of an early perceptual-cognitivist account of emotion. Gaita would certainly disagree with the cognitivist categorisation if it involves a picture of feelings as mere *causes* of cognitive error or disablement. Such accounts falsely assume, Gaita argues, that the cognitive can be specified independently of the fact of being moved, for some examples see (2000, xxxvi-xxxvii; 43-46; 2004, 33-36; 140).

sticks with us, goes some way in de-mystifying how certain terrible things we may have done are such that they resist being forgotten or fully submerged into memory. This, I take it, is a better way of capturing the idea Taylor was aiming at in her emphasis on the stain or mark of guilt. As I will discuss when elaborating Gaita on reparation, our vulnerability to remorse and the potential for it to endure should not be construed to mean that remorse is basically a form of self-indulgence or is excessively self-punishing. But, for the present, I will continue by elaborating the internal pole of remorse, what Gaita calls “the radical singularity of remorse”.

3.3.3 The inward trajectory of remorse: radical singularity

Until now, I have focused on the outward trajectory of Gaita’s account of remorse. It is distinctive of remorse, according to Gaita, that its steady focus is on the morally harmed party but this does not exclude the self in remorse. He calls the experience of self in remorse “radically singular”. But, given the importance and centrality of radical singularity, what exactly it amounts to requires some explanation. To do so, I will first consider the relevance of the first-personal or agential perspective. As we shall see, this is not quite sufficient to capture Gaita’s claim that remorse is radically singular, but it is part of the story.

3.3.3.1 The first-personal perspective

Given that radical singularity is descriptive of the agent’s experience in remorse, I begin by considering what Gaita says regarding the agential or first-personal perspective. There are several places where he discusses the agential or first-personal perspective of remorse, but I will only focus on one. It arises in his discussion of the neglect of the moral significance of “I did it” in consequentialist theories.⁴⁹ Specifically, he argues against consequentialists conferring exclusive moral significance to states of affairs that are the results of actions taken under moral consideration (2004, 54). Under this view, a

⁴⁹ The other significant location is in the example, discussed earlier in the chapter, of the Dutch woman’s statement that she hated Hitler for much, but above all, for making her a murderess (Gaita 2004, 47).

wrongdoer's sense that she has wronged her victim stems, then, strictly from the independent assessment of the value of the states of affairs relative to whatever standard of goodness is upheld by the particular theory (e.g., happiness, well-being, preference satisfaction, etc.). That a wrongdoer's sense that she has wronged her victim depends entirely on the independent assessment of the value of the states of affairs means that, any personal connection the wrongdoer may feel in response to their action, and its victim(s), is unimportant, morally speaking. The relevance of victims entering only with respect to the independent states of affairs produced also reflects the consequentialist version of Williams' objection against the picture of guilt that is either wholly detached from its victims, or, which sees their importance mediated by the theory's morally salient object (e.g., the principle or state of affairs). Namely, the loss of the special connection with the victims Williams takes to be what makes guilt morally distinctive and important.

But, the victims are not the only ones who disappear, for, consequentialist theories tend to ignore the first-personal aspect of morality.⁵⁰ In this connection, Gaita discusses Williams' first-personal form of regret, "agent-regret", which Williams contrasts with impersonal regret. Though, as I will argue in Chapter 4, Williams introduces agent-regret for more than one critical reason, central to its introduction is the aim of arguing against the tendency in moral philosophy to ignore the moral significance of the first-personal agential perspective, e.g., as expressed by "I did it". Gaita agrees with Williams, charging consequentialist theories with an inadequate sense of evil done, i.e., that it can matter to the agent, from her perspective that *she* caused the suffering. This is the moral idea at work in Gaita's interpretation of the Dutch woman who hated Hitler for making a murderess of her.

However, Gaita goes farther by also charging consequentialist theories with an inadequate sense of evil suffered. That is, it can matter from the perspective of those wronged that it was *she* who caused the suffering (2004, 55-61). This helps to capture the idea that, it can matter to me, morally speaking, that it was my mother who betrayed my secret, and not my

⁵⁰ As we saw in Chapter 1, impersonality is a key feature of the morality system, and it is one feature in which consequentialist theories are members of the morality system.

colleague.⁵¹ Although Gaita agrees with Williams' emphasis on the importance of the first-personal, agential perspective, he claims that, "The 'I' that is acknowledged in a serious response of agent-regret is not the radically singular 'I' that is discovered in remorse" (2004, 55). For Gaita, the "I" of "I did it", which is contrasted with an impersonal emphasis on agency and action in moral theory, is certainly an important feature of moral experience that Williams has rightly argued for. However, it is not the same thing as the radical singularity of remorse. Rather, Gaita is after something more specific than the distinction between the personal and the impersonal. To gain a better understanding of what exactly the radical singularity of remorse *is*, I will now turn to analyse the contrast Gaita presents between the "I" that is naturally subsumed into fellowship, and the "I" in remorse that responds to the recognition of guilt.

3.3.3.2 Common fellowship and radical singularity

To understand the "I" that is naturally subsumed into the "fellowship of we", I will first outline the "fellowship of we". The "fellowship of we"—the "we" of "we of this family" or "we of this circle of friends"—is, very basically, our sense of common human fellowship. It can be expressed at the level of our more or less immediate groups, all the way up to the more general level of our humanity (2004, 48). Gaita distinguishes between a sense of "we" that is of fellowship, and a sense of "we" that is enumerative. The enumerative "we" picks out a group based on a shared common characteristic, for example, as mortals, we are all going to die. The "we" here expresses the fact that all human beings, as mortals, will die. But, Gaita claims, the "we" can also carry a sense of the "we" of fellowship. That is, there is also a reading of "we", which invokes a sense that, *together*, we human beings share the same end.

According to Gaita, this sense of human fellowship is conditioned by the psychological phenomena human beings share, e.g., fear, anger, jealousy, and corrupt guilt feelings. Human beings can share, discuss and compare these feelings and, in doing so, we can express our fellowship. This common

⁵¹ Given his ubiquitous emphasis on the significance of partiality to moral reasons, for example, I doubt that, at this level, Williams would actively disagree with the idea that such theories have an inadequate sense of evil suffered.

fellowship has the power to console, which arises in connection with the thought that others too have done, and felt similar or the same things. Therapy groups for those suffering from loss, depression, diseases or addiction, for example, function on the basis of this fact and that, in explicitly sharing our sufferings with others, we can find solace or comfort, and possibly heal. This appeals to the old idiom that a problem shared is a problem halved. There is the obvious Humean point to be made here, namely that, this form of fellowship presupposes not merely the fact that human beings have the same or similar feelings but, more specifically, that human beings have the ability to *sympathise* or *empathise* with one another. Gaita acknowledges this when he writes that, our sense of fellowship is conditioned by “a common sympathetic responsiveness to the significance of ordinary human sufferings and joys” (2004, 217).⁵²

Human beings also, Gaita argues, seek solace and consolation in fellowship for their *moral* sufferings. Characteristically, we seek to assuage the painful sufferings of guilt, regret or shame by appeal to the idea that others have done the same thing or, that it is part of the human condition to err in the relevant ways. My feelings of shame after lying to my friend, for example, are consoled by the thought that my friend has lied to me too or, more generally, that everyone lies sometimes. Another helpful example is offered by Fricker, “The realization, for instance, that one is not alone in falling short of a certain level of courage may well comfort one and stop one beating oneself up about it” (2009, 318). As I interpret fellowship in its moral connections, when suffering guilt, shame or regret, the “I of fellowship” is responsive to the fact that others have done what I have done, or that others have fallen short in the ways I have.

The appeal to common fellowship or humanity, to sharing and unburdening in these thoughts, can offer consolation or comfort when feeling shame or regret. This comfort, in connection with shame or regret, need not function first-personally as exculpatory or as self-deceiving. That is, it need not be a form of moral avoidance or a corruption of one’s shame or regret to be comforted by the thought that others have done the same thing. It may, as

⁵² According to Gaita, “at the centre of all naturalisms”, is the thought that our sense of fellowship is conditioned by a common sympathetic responsiveness to others, and to the significance of human suffering and joy (2004, 217).

Fricker's example makes available, function to soften the severity of one's disappointment in oneself without necessarily diminishing the honest recognition that one was cowardly.

According to Gaita, however, remorse is distinguished from these other moral emotions in their different relations to the consoling thoughts of fellowship. As mentioned at the beginning of this section, Gaita contrasts the "I of fellowship" with the radically singular "I" of remorse. What this means is that, "Someone who is true to her remorse will always reject, as inappropriate, consolation that is based on her recognition of the guilt of others" (2004, 47). Where we naturally seek consolation in the sharing and recognition of the shame or regret of others, remorse, when genuine, does not. Internal to genuine remorse, according to Gaita, is the impossibility of the kind of consolation typically available in our sense of shared fellowship (which is not to say there is no consolation to be found in remorse).⁵³ It is this impossibility of consolation in sharing, in fellowship that centrally characterises the radical singularity of remorse.

The example of the Dutchwoman is instructive. Upon discovering that the three Jews were sent off to concentration camps and killed, the Dutch woman says "I have become a murderess". Following Gaita, if her remorse is lucid and serious, the Dutch woman cannot situate her suffering recognition that she is a murderer in relation to others, and their own guilty sufferings:

For what is sticking with the Dutch woman is not so much personal suffering as that specific kind of suffering that is remorse. Whereas other kinds of personal suffering can be consoled and alleviated by an awareness of the suffering of others, by a solidarity with them, guilty suffering or remorse (the feeling of guilt that is properly focused on the harmed party) can never be consoled in this way (Fricker 2009, 318).

The strict demands of remorse are such that her guilty suffering will not find comfort, dilution or reduction through being shared with others.

The impossibility of consolation in remorse through sharing leads Gaita to describe remorse as "a kind of dying to the world" (2004, 48). He writes:

⁵³ Gaita also claims that the demandingness of remorse extends to the impossibility of consolation by placing what one has done into a narrative (2004, 47).

I mean by it here what we mean when we say the world has become lost to those who are self-absorbed, as may happen to someone who is self-absorbed in the fear of her impending death. We mean that she cannot speak out of a sense of fellowship that is conditioned by unselfcentred and sympathetic responsiveness to others. The 'world' in this sense is a common world, and its kind of commonness is marked by the 'we' of fellowship. The radically and continuously self-absorbed lose, but are not thereby lost to, the common world (2004, 48).

Self-absorption and remorse are both forms of dying to the world, where "world" is understood as conditioned by fellowship. As forms of dying to the world self-absorption and remorse are both radically singular. As radically singular, in each, one is prevented from speaking "out of a sense of fellowship that is conditioned by unselfcentred and sympathetic responsiveness to others". This seems correct to say about self-absorption, but what does it mean in connection with remorse?

If remorse is distinctively characterised by its radical singularity, and radical singularity is to be understood as a form of self-absorption, then this seems plainly incompatible with Gaita's other distinctive characteristic of remorse, namely, its steady attention on the morally harmed party. But, this objection would be too quick, for it misconstrues the nature of radical singularity in connection with remorse. The first thing to note is that, for Gaita, remorse is not self-absorption (2004, 48). More specifically, Gaita takes the radical singularity of remorse to be expressive of "a dimension of ourselves that cannot enter into common and consoling fellowship with others" but that is not a form of self-absorption or self-centredness (2004, 49). The dimension of ourselves that we discover in remorse, I take it, is supposed to reflect the idea that, when we do confront our misdeeds, we confront them alone. With an existential overtone, the "something more than the first-personal agential perspective" Gaita is aiming to capture, is the sense of aloneness in this confrontation; of being the sole author of that action, and that no one else can share it, and so reduce or displace our culpability or sense of the significance of what we have done. It is important to note that what is revealed in Gaita's discussion are two aspects of the first-personal in remorse: the first-personal aspect (so often emphasised in Williams) that comes from the personal nature

of an individual's values, and there is the first-personal aspect of the *unsharability* of remorse. However, having this clarity regarding what radical singularity amounts to, unfortunately, does not help to resolve the question of the next section: how the two trajectories of remorse are related.

3.3.4 How are the two trajectories in remorse related?

On the one hand, there is the experience of the self in remorse as radically singular, and, on the other hand, an unwavering attention to the harmed party. At no point does Gaita address how these inward and outward trajectories or poles may be related. They remain, in his discussion, oddly apart. On one level, I take it, that there is no obvious tension in the idea that the position of the self in remorse as radically singular that is *not* a form of self-absorption or self-centredness, can attend to the harmed party in the manner Gaita takes to be necessary for remorse to be genuine. Radical singularity is not a claim about the direction of one's attention in remorse, which means that it is compatible with pointing to the central focus of one's attention in genuine remorse on those harmed. In fact, on Gaita's account, it is only from a position of radical singularity that one can truly attend to the reality of the party one has harmed.

Even if one does not accept that idea, I take Gaita to have provided a welcome antidote to the position taken by Taylor, which presupposed that, for the self to be present in an emotion, it is present only in self-regarding terms. The idea of remorse as radically singular allows for the presence of the self, which wholly other-regarding conceptions of remorse can neglect, whilst not falling into the mistake of assuming that, if the self is present, then it must be self-regarding. We might say that the "I" is not at the forefront of radical singularity. Namely, as I interpret Gaita, the emphasis in radical singularity is not so much on "I did it"; this merely registers that this is what I have done, and I am alone in relation to it. Rather, given the steady focus on the harmed party, the emphasis is likely slightly more on the "it", i.e., the harm done.

This suggestion of the relational nature of remorse goes some way towards our aim of seeing how the opposing trajectories in remorse might be

resolved. But, we are not there yet. We are still left with the basic puzzle: radical singularity prevents one from speaking out of a sense of fellowship or humanity, and yet, somehow, radical singularity must remain responsive to one's shared humanity otherwise it would simply collapse into self-absorption—the very picture of remorse we found in Taylor, and which we have been hoping to get away from. What is the relation between those two aspects of the self as they relate to others? Nowhere does Gaita consider this question, and it helps to reinforce the idea that he remains, in his own way, too stuck in considering the opposing trajectories of remorse independently of one another. This effectively repeats in microcosm Taylor's mistaken strategy—the dualism of self-regarding and other-regarding emotion. But, I have been urging that we can rescue Gaita's insightful characterisations of the two opposing forces within remorse by casting remorse overall as a relational moral emotion. On this picture, remorse is guilt-feeling of a kind that focuses on what *I* have done to *you*. Remorse does exhibit at once an outward and inward trajectory, and this is because remorse relates the subject and object of a wrongful action. This allows us to conserve his insights without repeating the mistaken dualism we hoped to avoid. With this discussion of both trajectories of remorse in place, I can now turn to considering another helpful aspect of Gaita's account: the corruptions of remorse.

3.4 Corruptions of remorse

The previous section set out much of the background needed to make sense of the many ways in which remorse can be corrupted. It is clear, from the discussion above, that seeking consolation in sharing in one's guilt would tend to corrupt remorse, for it would amount to a shirking of clear-eyed responsibility. Charting some of the typical pathologies of remorse is critical to Gaita's account because it serves to make his picture of remorse clearer, as well as, illuminating the demandingness he takes to be internal to genuine remorse.

However, there is another important reason for his focus on the corruptions or pathologies of remorse: "Contemporary hostility to remorse has been fuelled by a vivid sense of its corruptions, and we have, like Nietzsche who

brilliantly (if venomously) diagnosed the pathologies of remorse, failed to see that they are, after all, corruptions”(Gaita 2000, 34). By contrasting the radical singularity of remorse with the “I” of self-absorption, *and* with the “I” of fellowship, Gaita is in a position to diagnose what counts as corruptions of remorse, and thereby to argue against what he takes to be an exacerbating source of the contemporary “hostility” to remorse in moral philosophy.

The pole of remorse that concerns “what I have become” means that self-absorption is, as I shall elaborate in this section, an endemic risk in remorse. The proximity Gaita pointed to between the radical singularity of remorse as a kind of dying to the world, and the kind of dying to the world involved in self-absorption, gave a powerful indication of the central form that corruptions of remorse take. “Corrupt forms of remorse *are* a form of self-absorption” (2004, 49). Self-absorption, as a preoccupation with one’s own emotions, interests and/or situation, is, in Gaita’s terms, a kind of dying to the world, marked by an inability to “speak out of a sense of fellowship that is conditioned by unselfcentred and sympathetic responsiveness to others” (2004, 48). Some examples of expressions of self-absorption that one may be susceptible to in feeling guilt, include: those which turn on destructive forms of self-absorption like self-hatred, self-abasement, self-abnegation, or less self-destructive forms of self-absorption such as self-pity or self-consolation. Gaita makes an interesting diagnosis of some of these corruptions, “the ones that focus on a destructive sense of self-hatred and worthlessness—are often the result of confusing guilt and shame” (2000, 34). Similar to Bernard Williams’ account of shame and guilt in *SN*, Gaita takes remorse or guilt to focus on what one has done, whereas he takes the moral focus of shame to be on what is revealed about one’s character in having done that thing (2000, 34).

If suffering from feelings of guilt for something terrible one has done, there is the permanent risk that one’s attention will turn inward to an exclusive focus one’s own suffering. What happens, Gaita explains, is that the wrongdoer becomes absorbed in their own suffering in light of what they have become, and, in that exclusive self-focus, cannot turn their attention to those whom they have wronged in the manner appropriate to genuine remorse. “Then the ‘I’ of self-absorption becomes a false semblance of the radically singular ‘I’ who is

discovered in genuine remorse” (2004, 49). What can happen in such cases is that the “I” of self-absorption becomes a false semblance of the radically singular “I” revealed in genuine remorse, so that one takes oneself to be experiencing genuine remorse when one is not.

This falsity may be revealed, for example, in the character of one’s attention to those wronged. From this claim, I take it that we can derive a picture of what a simulacrum of genuine remorse might look like. The possibility of revealing a false semblance of genuine remorse can go some way in further diagnosing a central error in Taylor’s account: she mistakes the radically singular “I” for the “I” of self-absorption. As I argued in Chapter 2, Taylor assumes that whenever there is any kind of concern or attention to the self in feeling guilt, it must be self-involved. Thanks to Gaita’s account, what I am now in a position to say against Taylor’s view is that, though there is an endemic risk of collapse into self-absorption in remorse or guilt, it is a risk, and one that defines a pathology of remorse, and thus which is precisely to be contrasted with remorse proper.

Also following from the previous discussion, it is clear that seeking solace for one’s suffering guilt in sharing is a corrupt form of remorse. “Remorse is the only form of suffering that cannot legitimately seek comfort in a community of the guilty” (Gaita 2000, 33). This is one of the central insights of Gaita’s account: remorse that one attempts to dilute, reduce or divert by communing with others, (“it can’t be so bad if others have done it too”) is a piece of moral self-deception and/or a form of cowardice. It is a failure to accept one’s responsibility. A wife who, for example, is cheating on her husband might console the suffering guilt she feels for her betrayal by placing it in the familiar story that “everyone I know has cheated”. Dispensing of one’s suffering guilt in such a way is grounded in the idea that everyone is guilty, and so no one is (2004, 47). The familiar story, in other words, functions as exculpatory, and serves to avoid acknowledging her responsibility.

The story that all her friends have cheated is the wife’s response to what she has done, but what can this tell us about the corruption? A typical thought in this kind of case is that the wife is using the story to deceive herself, to protect herself from the attendant suffering of fully acknowledging the harm she has

caused her husband and what she has become. But the corruption of her guilt, on Gaita's account, is not simply the outcome of self-deceit or an inability to understand the nature of the harm of her betrayal. Her corrupt response is a *form of her failure to understand* what she has done (2004, 35). This reveals another critical aspect of Gaita's account that I will take up in the next section: he takes moral response and moral understanding to be interdependent.

3.4.1 Are all forms of corruption in remorse forms of self-absorption?

Aside from seeking solace in sharing, Gaita also identifies, as a corruption of remorse, the hope that, with time, remorse can be healed (2000, 33). Namely, one cannot be consoled in genuine remorse with the mere passage of time. Here, once again, is the idea of the wrongdoer avoiding the confrontation with the full significance of what she has done. But, is this kind of corruption a form of self-absorption? This question of whether all corruptions of remorse are forms of self-absorption arises in light of Gaita's claim that corrupt forms of remorse are forms of self-absorption. There are strong reasons to understand this claim as meaning that, *all* forms of corruption in remorse are expressions of self-absorption. In a comparison with grief Gaita writes,

Remorse is a recognition of the reality of another through the shock of wronging her, just as grief is a recognition of the reality of another through the shock of losing her. Both are liable to egocentric corruptions. Our dependencies, even at their best, tread a fine line between awakening a sense of the reality of another and submerging that sense in one of the many forms of egocentric absorption. Exactly the same is true of remorse. But the egocentricity is not merely a feature of the corruption, it is its central feature (2004, 52-53).

Given this, I think one can make the claim that Gaita takes all forms of corruption in remorse to be variously forms of self-absorption or egocentricity. I take it that, what makes all forms of corruption in remorse forms of self-absorption is the preference or primacy the wrongdoer gives to the pursuit of the assuagement of their own painful, guilty feelings. This primacy of one's feelings over and above those harmed can take various forms, including the idea that if one just gives it time, for example, all will be fine.

I think this leaves the question of how the corruption of seeking consolation in fellowship is to be understood as a form of self-absorption. How does the “I” of self-absorption seek consolation in sharing? This is particularly interesting given that one’s sense of fellowship is conditioned by an *unselfcentred* sympathetic responsiveness to others. The answer I think Gaita could give is that, in seeking solace from fellowship, the wrongdoer is aiming to, at least partially, erase the wrong from her view of herself. This self-serving self-deception prioritises self-image over the reality of the harmed other, and is in that sense a form of self-absorption.

Before moving on to the next section, it is worth remarking on the connections between the corruptions of remorse, and one of the main reasons Gaita takes there to be hostility to remorse in moral philosophy. The reasons for the hostility to remorse are varied, ranging from “a reductive functionalism about value, that focuses on the superficial thought that guilt serves no purpose (why should it?), to the most high-minded of them, which scorns remorse as a form of self-indulgence at the expense of a proper concern for the victim of our wrongdoing” (2004, 51). On this first reason for the hostility to remorse, Gaita, as we shall see in the next section, is concerned to argue that remorse has purpose, particularly as it relates to moral understanding for both self and others. But, it is the latter reason, taking remorse to be a form of self-indulgence, that has been addressed in this section. Aside from the fact that Taylor’s account of guilt is subject to the objection of self-indulgence, and the ethical charge of it coming at the expense of a proper concern for the victim, Gaita’s account is not. An important upshot I take from Gaita’s work is that, in addressing and explaining the corruptions of remorse, he reveals its internal demandingness. But, more crucially, he also shows that although there is an endemic risk in remorse to self-absorption, self-indulgent remorse is a form of *corrupt* remorse.

3.5 Remorse and moral understanding

Returning to the case of the cheating wife, her response to what she had done was taken to reflect her lack of understanding of the significance of her betrayal. If the wife had properly understood what she had done, the thought

goes, she would have discovered what it is to be someone who betrays another by cheating. Indissolubly linked to this discovery would have been, not only the realisation of the reality of her husband as irreplaceable, but, also, the harm she caused herself in her betrayal. In other words, had she properly understood what she had done, she would have discovered what she had become in light of what she did. According to Gaita, the realisation of what one has become in light of what one has done “is a necessary condition of the recognition of the evil we have done” (2004, 63). If the wife did not recognise that she had become the particular kind of wrongdoer that is a cheater, we could not say that she understood what she had done. As such, Gaita would say that her remorse was corrupt but, equally, that she did not have a proper understanding of the wrong she committed. It seems that without the moral understanding of the evil done, one is not able to respond with genuine remorse. In the same way, if one does not show genuine remorse one cannot claim to have the proper moral understanding. This relation, following Gaita, explains the constitutive relation between remorse and one’s understanding of wrongdoing, thereby situating remorse as an essential part of moral understanding.

The discussion is furthered when Gaita considers an objection to his account namely, “that remorse is a reaction to an independently intelligible conception of wrongdoing and is a psychological state structured by it” (2004, 55). The objection captures the idea, familiar from cognitivist accounts, that we cannot feel remorse without judging that what we did was wrong. Certainly, Gaita’s account of remorse recognises that one must judge that what one did was wrong, but he wants to deny that this is independently intelligible from the feeling of remorse. This speaks to Gaita’s commitment to the idea of the inseparability of thought and feeling in remorse. Appealing to Wittgenstein’s idea that “pity is a form of the conviction that another is in pain”, Gaita wants to say that “remorse is a form of the conviction that we have done something wrong”.⁵⁴ Namely, our feeling of remorse, and our sense that we have done

⁵⁴ See *Philosophical Investigations* §281, where I take Wittgenstein to be arguing that we should not focus on the question of what condition (inner or outer) the word “pain” might refer to, but rather pay attention to the role of expressions and attributions of pain in our lives, for the way the use of the word “pain” is bound up with our ways of responding to, and dealing with, others (see McGinn 2013, 143-170).

something wrong are conceptually interdependent; in rough analogical paraphrase of Wittgenstein, “our natural dispositions to remorse are one of the determinants of our concept of wrongdoing” (Gaita 2004, 55).

Given that our feeling of remorse and our sense that we have done something wrong have a constitutive relation, it follows that there is a *relation of intelligibility* between remorse and wrongdoing. Gaita writes,

I am not saying that we can tell what is evil and what is not by considering whether we feel remorse for what we do...[it is that] we cannot say that something is morally wrong unless it is an intelligible object of remorse...Our understanding of remorse is not conceptually recessive in relation to an independently intelligible conception of moral wrongdoing. The two are, at least, equal partners (2004, 60).

Here we find one of Gaita’s strongest insights: we cannot *say* that something is morally wrong unless it is an intelligible object of remorse. Intelligible remorse, in other words, sets the parameters for what can possibly count as morally wrong.⁵⁵ Clearly, this discussion presents a very different picture from the standard cognitivist picture presented in the objection (which does not take away from the characterisation of Gaita as offering a cognitivist account). Where the objection situates remorse as a secondary reaction relatively inconsequential to our understanding of the wrongdoing, Gaita is arguing that remorse is in fact central to our understanding of the wrongdoing. Therefore, remorse, according to Gaita, is central to moral understanding.

It is important to emphasise that Gaita is *not* claiming that simply because someone feels remorse in response to something done that this is sufficient to render their response rational or appropriate with respect to wrongdoing. On the circle of the constitutive interdependence of our responses and our concepts, and their identification of the objects of our responses, Gaita writes,

⁵⁵ From this truth, according to Gaita, we have a test for moral theories that aim to deliver the claims of “Reason”. If such moral theories consider something as intelligibly wrong or evil when we would not say that someone ought to feel remorse for doing it, then we can judge the moral reasoning employed by the theory to be wrong.

The responses that form and are formed by our sense of belonging to a common kind cannot be elicited by beings that do not look and behave like us. This is...because those responses are built into the concepts with which we identify what could be appropriate objects of our responses. This is a circle—a non-vicious circle—from which we cannot escape without losing the relevant concepts ...We have to see how we respond. And reflect on our responses of course. But such reflection cannot escape the circle made by the interdependence of such responses and the concepts with which we identify their appropriate objects (2000, 269).

Owing to his acceptance of the facts of cultural difference, and cultural change over time, Gaita can acknowledge that, at a given moment in time, what someone feels remorse over could be something that has yet to be recognised in a particular culture or community as being an intelligible object of wrongdoing (2004, 60). Accepting this circle between remorseful responses that condition our concept of wrongdoing, and our concept of wrongdoing as identifying the intelligible objects of our remorse, is unproblematic because, as Gaita explains, we can step back from our responses in reflection. This is why the circle is characterised as *non-vicious*.

However, as the final line in the passage above makes clear, there are limits. “[A]lthough our responses and reactions are not self-authenticating, the interdependence between some of them and what they are responses to places limits on the degree we can step back from our responses in order to judge whether they are rationally appropriate to their objects” (2004, 118). The constitutive interdependence of our moral responses and our moral concepts does present an internalism, and imposes limits to how far we can intelligibly step away from our responses in reflection in order to judge whether they are rationally appropriate to their objects. But, this is not a real problem, because we are not locked-in to the response-concept circle. That there are limits on how far we can step back does not short-circuit the fact that we have the critical resources, for example, in our critical concepts, that help us to determine whether something is falsely sentimental, dishonest or cowardly, and to reflect on, and question, our responses and their objects. This way of understanding our responses and their objects will take on greater significance in the next section, as Gaita takes aim at the standard view in moral theory that remorse, as

a first-personal expression of moral responsibility, is rationally restricted to the voluntary.

3.6 Blame, moral responsibility and remorse

As remorse is concerned with a wrong one has committed, questions of second- and third-personal responses to the wrong one committed naturally arise. Specifically, Gaita considers blame, and its characteristic connections with moral responsibility. Standardly, this is the view that the domains of blameworthy wrongdoing and morally responsible wrongdoing are coextensive. According to Gaita, such a view maintains that blame can only be appropriate and/or rational if the wrongdoer can be said to be morally responsible for what they have done, and, one can only be morally responsible for what one has done if it was done voluntarily or intentionally.

Of course, this is the familiar position outlined in Chapter 1 on the morality system⁵⁶, and it is powerfully driven by the view Williams called “the doctrine of the voluntary”. That is, it would be *unjust* to blame someone for something they have not done voluntarily or intentionally. Moreover, as we saw in my construction of MS-remorse, on a view that treats the domains of blame and moral responsibility as coextensive, it is typically considered inappropriate or irrational for the wrongdoer to express remorse for an action for which they are not blameworthy/culpable. Again, as I explained in that connection, remorse is characteristically taken as self-blame—the first-personal mirror of second and third-personal blame—and so is assumed to be rationally restricted to the voluntary. It is, on such a view, irrational to feel remorse for involuntary actions.

As we saw in Chapter 2, Taylor implicitly argues, with her accommodation of instances of bad moral luck, against the mirroring of the conditions of blame in guilt. As I shall argue in the next chapter, Williams, despite a popular misinterpretation, also argues explicitly against this picture. It is not a coincidence, I take it, that, in both cases, the first-personal perspective is taken as the starting point, and, implicitly, as bearing a certain kind of authority.

⁵⁶ As Paul Russell remarks, “The narrow construal of responsibility requires that we understand moral responsibility within the conceptual resources provided by the morality system, making notions of obligation, wrongness, and blame essential to the analysis of moral responsibility” (2013, 185).

As mentioned in the first section, Gaita also takes this route. In his line of argument, he takes explicit aim at the standard view of the domain of moral responsibility as coextensive with the domain of potential blame by arguing against the idea that remorse is restricted to the voluntary. His central target is the idea that we cannot intelligibly feel remorse for bad actions, or bad consequences of actions, for which we are not blameworthy.

A brief return to the case of the Dutch woman is helpful. On the morality system account, there is nothing for which the Dutch woman can be blamed. So, though it is understandable on a psychological level, that she might feel remorse, her remorse is in fact irrational as she is not culpable for the murder of the three Jews. It was not her fault they were murdered, and so her remorse is irrational. Taking the lucidity of the Dutch woman's response to reflect the authenticity of her object, what Gaita is concerned to argue is that it need not follow from a person's non-culpability that they are not morally responsible, or that remorse would be an irrational response to what they have done.

For Gaita, there is a distorting tendency in moral philosophy to attach moral responsibility too tightly with culpability. His reply is to argue that culpability is better understood as a narrower conception operating within a broader conception of moral responsibility. To explore Gaita's arguments against this narrow conception of moral responsibility, and his alternatively broad conception of moral responsibility, in what follows I will first outline mistakes of the narrow conception. Once these mistakes have been set out, I present Gaita's discussion of the canonical example of Oedipus as an argument for the wider scope of moral judgement. Finally, drawing on the non-vicious circle of responses and concepts presented in the previous section, I outline Gaita's broad conception of moral responsibility.

3.6.1 Some mistakes of the narrow conception of moral responsibility

Before elaborating Gaita's broader conception, it is important to note, particularly in its connections with Williams' critique of the purified conception of the voluntary, that Gaita's critique is not simply critical of the assumption that moral responsibility is tied too closely with potential culpability. Gaita claims

that not recognising other forms of moral responsibility can result in one or more of the following mistakes, as Fricker helpfully explains:

Those concerned to apply some social understanding to the phenomenon of wrongdoing are led into a kind of social determinism that underplays freedom and effectively lets the culprit off the moral hook; and, in turn, this undue moral neutrality can inspire those who are concerned to apply a moral vocabulary to exaggerate the element of choice so that they wind up espousing an ‘implausible voluntarism’ that depicts everyone as simply free to choose what they do (2009, 316).

This interpretation captures a picture manifest in both a deterministic conception of the will, and a freedom conception of the will—arguably the same picture of the will in action traced in Williams’ critique of the purified conception of the voluntary.

Under the socially determined conception, the individual cannot be held responsible for any of their acts and, therefore, cannot be culpable or blameworthy. The character of this picture encourages the extinguishing of *any* moral judgement; again, we encounter the assumption that without blame there can be no moral judgement. Under the freedom conception, every individual is conceived of as infinitely free or unconstrained to choose what they do, and are thereby culpable or blameworthy for every “wrong” choice they make. But Gaita’s account accommodates the idea that one can do many things voluntarily (under a certain description) without realising *the significance* of what one has done.

As we shall see below, there is no question that Oedipus intended to kill the king. The issue is that Oedipus did not realise the full significance of the killing because he did not know it was his father that he’d murdered. Relatedly, as the Dutch woman’s case brings out, we can be responsible for things we freely or voluntarily do under that description without being the *direct* cause. “Her moral response is clear-eyed and her sense of guilt is about her proximal role in the causal chain that led to three murders” (Fricker 2009, 317). Here, Gaita stretches the domain of responsibility even farther than Taylor, for, in my elaboration of her account, we saw that she relies on the notion of being the immediate or direct cause of the morally bad thing.

Like Williams, Gaita takes the mistake of an implausible voluntarism to contain the moralistic distortion of a restricted conception of moral responsibility that only recognises blame as a second- or third-personal response to wrongdoing. But, it is not just any conception of blame this restricted picture of moral responsibility encourages—it contains a style of blame he associates with what we now call “judgementalism”—a finger-wagging style of blame. Gaita writes,

“[[J]udgementalism—judgement that would *blame* her (bearing in mind all the connotations of that word), that would encourage one to point a finger at her and to turn one’s back on her. But a preparedness to see (and in that sense to judge) a situation in a severe moral light, while at the same time refusing to blame, strikes some people as incoherent. That, I think, is the effect of a moralistic conception of morality” (2000, xvii).⁵⁷

Here, Gaita is arguing for the possibility of serious moral judgement of what someone else has done without judgemental blame. Moreover, Gaita takes it to be a moralistic conception of moral responsibility (and of morality) to assume that it is a necessary feature of serious moral judgement that it take the form of blame, finger-wagging style or not. For Gaita, to not see that we can be held responsible, i.e., *answerable to what we have done and its significance*, without blame, is to be stuck in a moralistic conception of moral responsibility (and morality more generally).

3.6.1.1 *The case of Oedipus: broadening moral judgement from narrow blame*

In another connection with Williams, Gaita turns to the canonical example of Oedipus to illuminate the possibility of serious moral judgement without blame. He writes,

⁵⁷ In very similar terms to Williams’ diagnosis of the morality system, on the origins of the moralistic conception of morality Gaita writes: “It was generated, on the one hand, by a legalistic intrusion into morality of categories of culpability necessary for us to account for the wrongs we do one another in the political realm, and on the other hand, by an unsavoury tendency, that appears to go as deep in us as our receptivity to free-floating malice, to be judgmental” (2011, 161).

In Oedipus Rex, Oedipus kills his father and marries his mother. He does both deeds unintentionally because he does not know that the man he killed was his father and that the woman he married was his mother, and he is not culpable for his ignorance. When he discovers what he has done, the language of Sophocles' play leaves us in no doubt that his response is remorse, and it also leaves us in no doubt that the response of the chorus is pity. Its pity has many layers but is, at its core, pity for the wrongdoer that Oedipus, entirely blamelessly, has become. It is a severe pity, however, because it will not let him evade the fact—as he tries to in Oedipus at Colonus—that, blameless though he is, he is a wrongdoer and 'miserable and pitiable' just on that account (2011, 160).⁵⁸

This passage is doing work on several fronts. Firstly, the example of Oedipus is a clear case of an expression of remorse where conditions of culpability do not hold (particularly under the description of, at the various times of action, not being in a position to know the *significance* of the actions). In this way, the example is aimed at supporting the earlier claims that (i) rational remorse, as an expression of first-personal moral responsibility, extends beyond the conditions of culpability and, thereby, that (ii) moral responsibility extends beyond conditions of culpability and blameworthiness.

The passage also serves to explain a form serious moral judgement may take that is not blame. This can take form in pity or feeling sorry for the wrongdoer, for what they have done, and for what they have become. Unlike judgemental blame, which does not register sympathy for the wrongdoer, in pity, as a response to the wrongdoer and their actions, there is the expression of a willingness to understand the wrongdoer. Pity or feeling sorry for the wrongdoer registers one's capacity for sympathy for the wrongdoer whilst holding them responsible or answerable to the severity or seriousness of what they have done. In this way, we can understand Gaita as pairing remorse, in these cases of bad moral luck, or what he might call "blameless wrongdoing", with the second- or third-personal responses of pity or feeling sorry for the one who brought about these events, however non-culpably.⁵⁹

⁵⁸ Also see Gaita (2000, 93-96; 2004, 44).

⁵⁹ Interestingly, Williams' remarks on Oedipus and pity are almost the same, "[the Chorus'] pity still acknowledges the presence of his past. It is aroused not just by what he later suffered, but *by what he did*, and *by his own acknowledgement of what he did*: how he sees what he did and how others see it form, as they must in any such case, a pair whose parts structure each other" [my emphasis] (1993a, 71). Williams also makes this responsive pairing between agent-regret and feeling sorry in his discussion of the lorry driver (1981b, 28).

Like Williams, Gaita situates his account within a Western historical context, one that acknowledges the possibility of tragedy and moral luck. He writes,

The terrible possibility that we might become wrongdoers though we are blameless marks one, historically constituted, pole of our (Western) moral sensibility. Oedipus Rex is perhaps its greatest representation. More commonly we may find ourselves, through no fault of our own, in a situation in which we wrong someone whatever we do and, therefore, inevitably become wrongdoers whatever we do. In such cases accepting responsibility does not mean accepting blame, but being lucidly responsive to the moral significance of what we have done. Remorse is the form of that lucid responsiveness (2011, 160-161).

In this passage not only does Gaita recognise cases of blameless wrongdoing but, also, the possibility of becoming a wrongdoer through moral dilemmas (including landing in a dilemma through no fault of one's own). Such cases, for Gaita, are not distinct from more culpable cases with respect to the appropriateness of remorse understood as an acceptance of one's responsibility, and as a responsiveness to the moral significance of what one has done. Remorse, as a first-personal expression of moral responsibility, is, therefore, broader than blame and the conditions of culpability. To not recognise this, to claim that "though it is understandable that Oedipus should be profoundly upset over the fact that he killed his father and married his mother, it is irrational for him to feel remorse because he is not culpable for either," is to fall prey to the moralistic distortion of morality (Gaita 2011, 161).

3.6.2 Gaita's broader conception of moral responsibility

Clearly, Gaita is mounting a case for a broader conception of moral responsibility. He provides the following formulation: "We can say that *a person is morally responsible for what may claim her and us in one of the many forms of serious and lucid moral response*" (2004, 45). Regardless of whether someone has done something wrong voluntarily or intentionally, that she can hold herself and/or be held by others in serious moral response for what she has done reflects her moral responsibility.

Unlike the standard view, under Gaita's conception, it is possible to acknowledge that someone is not culpable or to be blamed for what they did but, that they can nevertheless be held morally responsible for it. Gaita writes,

"To hold someone responsible in this sense means to hold them, to fix them, in a lucid response to the significance of what they did. It means that the moral significance of what they did must not be evaded, neither by them nor by us, but it does not, thereby, mean that we find fault with them, that we can accuse them, or that we find them culpable" (2004, 44).

But *what* might claim us in serious response? The upholder of the standard view might say, how can we judge the appropriateness of the response to the action? How can we tell if the response should be serious or not? This, as I shall explore below, is where Gaita's non-vicious circle of responses and concepts most obviously plays a critical role in his account of remorse.

On a traditional cognitivist account, we make the assessment of the action independently of the response to it, as saw in the elaboration of Taylor's account in Chapter 2. On this view, response occurs in relation to an independent moral fact that, once grasped, others can refer to in making their judgements regarding the appropriateness or rationality of the wrongdoer's response. That is, the response follows upon an independent critical assessment of the wrongdoing. Of course, as I elaborated in the previous section, this is not what Gaita wants to claim. Again, he writes "*what* can claim us in serious moral response need not, at all levels, be established independently of what is revealed to us by authentic and authoritative response"(2004, 45). The thought is that a genuinely lucid and serious response can tell us *what* counts as morally significant.

Reflecting the constitutive interdependence between moral responses and intelligible objects of moral response, this is not to say that our responses are not open to scrutiny. What it does break apart is the thought that the objects of our moral response are *at all levels* independently intelligible. But, what makes a response authoritative? What I take Gaita to be claiming is that, what makes a response authoritative is the quality of the response. Namely, whether the response is self-pitying or sentimental as opposed to serious and lucid.

These types of descriptions are part of what Gaita calls “the critical grammar that determines our sense of the authority” (2004, 45). The critical grammar regulates which moral responses we judge to be sentimental or cowardly or self-indulgent as opposed to the more authoritative responses, which we judge to be serious, lucid, or honest.

Gaita’s commitment to the idea that our inherited critical grammar regulates what we judge to be sentimental as opposed to serious suggests that he trusts our capacities to make sound judgements regarding the quality and authenticity of our responses. Of course, this is not to say that our determination of the quality of a response is not related to what we might say about the kind of thing a person did independently of the quality of the response (Gaita 2004, 46). So, that the Dutch woman’s response is recognisably serious, as opposed to sentimental, can teach me what the proper object of her response is (i.e., her non-culpable moral responsibility for the murder of those three Jews). But it is also instructive of the *moral significance* of the object. The point is that it was not from some neutral point of view that the moral significance of what the Dutch woman did became clear. Rather, it was the Dutch woman’s remorseful response that illuminated the moral significance of what she did.

If lucid remorse can illuminate the moral significance of what one has done, then we can understand how it is that second and third parties might even learn from such cases. Fricker offers a helpful summary, “the authenticity of the Dutch woman’s response can teach one the moral significance of what she is responding to. One can learn to see something one might have missed – the nature of her responsibility for what happened – by seeing it through her eyes”(2009, 316). Formulating the positive possibilities of remorseful response not simply in terms of the wrongdoer’s moral understanding, but also with respect to the moral understanding of second and third parties, functions as Gaita’s reply against a common objection that sees remorse as “unproductive” of positive moral change, e.g., moral improvement or development (2004, 189). There is also the often related objection that remorse is unproductive with respect to what it ought to be productive of, that is, reparation, which I now turn to in the next section.

3.7 Reparation

Surprisingly, Gaita has little to say on reparation. However, his stray remarks in combination with the preceding analysis offer materials with which to work up an interesting conception of reparation, or so I shall try to show. I will first elaborate his remarks on reparation as they appear in response to a misconception of remorse. Working from these remarks, and some materials previously elaborated, I will briefly sketch a picture of reparation that makes use of some of the key insights of Gaita's account of remorse.

3.7.1 A misconception of remorse

Gaita's remarks on reparation appear in his discussion of two common, and related, misconceptions of remorse. The first misconception, familiar from the foregoing discussion of cognitivism, charges that remorse cannot be fundamental to moral understanding because it is a mere feeling and, as a feeling, it is recessive to achieving a proper moral understanding of a wrong one has done. The second misconception, which I will focus on, relates to the idea that one learns about one's wrongdoing in reparation.

This second misconception assumes that "the recognition that we have wronged someone is best revealed in reparation, and that is where we should locate a proper sense of the seriousness of what we did. Reparation rather than remorse is expressive of what it is to take another seriously"(Gaita, 2004, 53). As I interpret Gaita, the problem with this misconception is that it assumes that what is most morally salient as a response to wrongdoing is reparation and, thereby, that the moral value of remorse is reducible to reparation. If the recognition that we have wronged someone is best revealed in other-directed reparation, then the feeling of remorse begins to look like it is not a serious other-regarding response. As I understand Gaita's point, by reducing the moral value of remorse to reparation, this misconception indirectly encourages the idea that remorse, as a mere feeling, must be a form of self-indulgence.

Gaita's reply is not to deny a deep interconnection between remorse and reparation: "Remorse and reparation are not exclusive of each other, and without a serious concern with reparation, where it is possible, remorse would

be corrupt” (2004, 53). If remorse without reparation (where possible) is a sign of corrupt remorse, then Gaita takes the relation between remorse and reparation to be deeply important. In order to make reparations, according to Gaita, one must understand *what* one is making them for. This serves as an objection to the second misconception: a willingness to make reparations depends on one’s understanding of what one is making reparations for. And, as Gaita has been arguing, remorse, as a form of moral understanding, is how we come to understand what we are making reparations for. Reparation, therefore, depends on our moral understanding as it is discovered in remorse. Consequently, making reparations, as the second misconception would have it, cannot be conceptually independent of feeling remorse.

3.7.2 Sketching reparation

Reparation is often an expression of remorse as part of feeling badly for what one has done. It is aimed at showing a desire to make things right, which can take many forms, and, often depends on what one has done. One might apologise in various ways and/or take action; either of which can be singular instances or part of a prolonged project of repair. There is also the possibility of symbolic or gestural reparation like kissing a child’s bruise, which cannot make the bruise disappear, but there is a sense in which it can still serve to make the situation better. Moreover, it can matter with respect to accepting reparation as genuine, what kind of “spirit” it is made in, for example, whether it seemed perfunctory, inattentive, or kind and warm.

Sometimes, what one has done is so terrible there is only the possibility of a ceaseless recognition of what one has done. A young man, for example, had a few too many drinks at his friend’s party and, on his way home, hits another car and kills both passengers. He feels remorse for what he has done, and apologises to the victims’ families. As is the case with killing and murder, one cannot repair the damage with the individual directly harmed but must direct one’s attention to those who are left behind. As part of the process of reparation he steadily builds a project that sees him working with the loved ones of those he killed to speak about drunk driving. Of course, sometimes, there is nothing

one can do or say. Much depends, I suspect, on how one conceptualises the nature of reparation, and whether it seeks to “undo” the harm as though the aim is to erase it or make it disappear, or, whether the harm is understood as something to be recovered from. Some of these ideas will find expression in my sketch below, though I think there is much more work to be done in this neglected area.

On Gaita’s account, it is clear that reparation, when it is possible, requires the moral understanding that comes about through genuine remorse. This moral understanding, for it to be full, must reflect each of the twin features of remorse: a sense of responsibility for what one did, and what our victims suffered, both of which are not reducible to each other or anything else (Gaita 2004, 54). But, again, if one’s remorse is genuine, then one’s moral understanding cannot be strictly in terms of the natural harms. Rather, the moral understanding revealed in genuine remorse involves grasping the harm of the moral significance of what the victim has suffered, and the moral significance of what it means to become a wrongdoer. It follows that appropriate reparation, as a response to one’s remorse, must thereby reflect one’s moral understanding not just on the level of natural harms but also in terms of the relevant moral significances. As I will explore below, the idea that remorse is a form of moral understanding in connection with this distinction affords some interesting materials to sketch the beginnings of an account of reparation. This account captures the idea, encountered in Taylor, where inadequate or incongruous or incommensurate reparation can act as a signal of corrupt (as Gaita would say) and/or undeveloped (as I will say) remorse.

Take the example of an old woman living alone who is violently robbed by a young man she once knew. If the young man now orders a new television to be delivered to her house to replace the one he stole, we might think that it is insufficient to the character and significance of what his actions meant for her. Why we might think it insufficient to the harm caused can be explained by Gaita’s distinction: the replacement of stolen goods does not address the significance of the betrayal as a separate and irreducible cause of the harm suffered. Moreover, the betrayal she suffers is not just at the level of a home invasion but, also, at an even more personal level, in the breaking of a once

trusting relationship. In this way, it would undoubtedly matter to the old woman in alleviating her suffering, and rebuilding her sense of trust that, if the young man were to try and repair the harm, he would do so in a spirit that recognises the significance of his betrayal *for her*. What this means for his reparation is that, for it to be effective with respect to bringing about her recovery, it cannot merely be expressive of a general understanding of the significance of betrayal. What it will have to express is what he imagines or comes to understand the betrayal to have meant for her, and, for his understanding to be conveyed to her.

As this example suggests, natural harms can be fairly easy to explain in connection with reparation. This is because natural harms are often straightforwardly compensated for by insurance money, the replacement of goods lost, payment of hospital or doctor's bills, legal fees, etc.⁶⁰ Of course, there will be cases where the natural harm cannot be compensated for in material or financial terms at all (which is not to say that victims are not owed restitution in such cases). There will also be cases where these forms of compensation will be satisfactory with respect to the victim's recovery. But, as the example also made clear, compensation of the natural harm cannot always speak to the victim's recovery by way of reparation. Part of why I take this to be the case is that, part of what it is to understand the moral significance of the harm the victim suffers, is not simply to understand its moral significance in general terms, but to understand the particular instantiation of it with respect to the individual harmed.

One of the upshots of Gaita's cognitivist conception of remorse is that we can point to specific epistemic sites of misapprehension in reparation. This is particularly the case in relation to the more difficult epistemic demand of grasping, at multiple levels, the moral significances of one's wrongdoing. That this places a high demand on the wrongdoer is further evident in what Gaita has to say about emotions as forms of understanding. Emotions as forms of understanding are very sophisticated, requiring a collection of sensitivities, for example, empathic imagination, moral understanding and open mindedness. Gaita follows this up by claiming that corrupt moral responses are like most

⁶⁰ This idea is also captured by Williams in his "insurance test", see (1981b, 29-30).

forms of moral corruptions in being a corruption of sensibility (2004, 35). These demands help to suggest some of the important ways in which reparation may fail and/or signal corruptions of remorse.

Against this backdrop, I want to exploit the helpful idea that we manifest our moral understanding in how we respond (i.e., when, where, how, to whom, etc.). To illuminate this point, I formulate what can be called a “reparation test”: if reparation is a signal of the quality or character of one’s remorse, then we can understand the quality of reparation (or its absence) as indicating the wrongdoer’s state of mind and their moral understanding. This can help us make an important moral distinction in cases where a person(s) does not demonstrate an understanding of the moral significances (and so can be seen to be not demonstrating remorse). On the one hand, a person can simply be not remorseful. On the other hand, a person who has guilty feelings, and although not trying to re-situate those feelings (which would signal a lack of genuine remorse), has yet to fully grasp the moral significance of what they have done, and what their victims have suffered.

This distinction, and particularly the second case I am entertaining, is a departure from Gaita as I am considering the possibility of a person who has yet to fully grasp the moral significances because, for example, she is too inexperienced or immature. She is approaching her wrongdoing, and attempting to make reparation in the right spirit, but nevertheless is not yet in a position to grasp the full moral significance of what she has done, and what her victims have suffered. She is not seeking consolation in sharing nor is she hoping that, in time, it will heal. It is simply the case that she is not yet in a position to fully grasp the moral significances. This is the everyday point that what understandings are available to the wrongdoer are going to make some truths more or less available to her. In other words, departing from Gaita, I think there is something morally significant in the idea of an imperfect but good-enough understanding in remorse. On this account, it follows that remorse need not require, as Gaita suggests, either the actuality or idea of the possibility of a perfect moral understanding. This idea of a good-enough understanding in remorse appeals to the idea of a process of discovery, whereby one’s moral understanding develops over time or through an accumulation of experiences.

This may seem to diverge from Gaita's idea of remorse as an awakening, but it need not. It could be that what Gaita calls "genuine remorse" is actually an ideal revelatory culmination of a process of discovery, which seeks to understand the moral significance of what I have done, and what the victims have suffered. The norm may often be imperfect moral understanding, but, so long as it is good enough, then the remorse will similarly be good enough.

3.7.3 The wrongdoer and forgiveness

An upshot of Gaita's account is that it makes clear that genuine reparation, where possible and proper, is conceived as essentially concerned with the recovery of the victim. As we shall see here, this recovery may take the form of forgiveness, where forgiveness is one of the central ways in which the burden of remorse or guilty suffering can be relieved. Summarising Gaita's position, Fricker writes,

Remorse sticks to one as a lone individual, and one must deal with it, perhaps recover from it, by other means than through the therapy of fellowship. What about time? Time heals most things, but not remorse: it may be half-forgotten, but in itself remorse will remain unchanged by the mere passage of time. The mystery of the phenomenon of remorse, of guilty suffering, is reflected in the mystery of what can bring recovery: repentance, atonement, forgiveness, and punishment (2009, 319).

Not unsurprisingly, reparation, for Gaita (unlike Taylor), is not a straightforward means of recovery for the person feeling remorse. Relief from the burden of remorse, as we saw in connection with the impossibility of sharing one's guilt and thereby halving it, cannot be lessened or diminished. What may bring about recovery for the wrongdoer is forgiveness, and, reparation, where possible, can, help bring it about.

What these connections between remorse, reparation and the power of forgiveness suggest is that my previous claim that remorse is best understood as a relational emotion indeed has force. As mentioned above, it would undoubtedly matter to the old woman in rebuilding her sense of trust that, if the young man were to try and repair the harm, he would do so in a manner and spirit which recognises the significance of his betrayal *for her*. And, in doing so,

it may be that the old woman forgives him, and that the young man comes to feel relief from the burden of his guilt. Following Gaita, it seems correct to say that reparation, by way of encouraging forgiveness, could have a positive effect on him as well as on her. This reinforces the idea that remorse is essentially relational, for it takes seriously the idea that restoration of imbalance through reparation essentially happens *between* the wrongdoer and the victim, so that the way they relate is transformed first through remorse and any associated reparations, and then through the forgiveness this may prompt.

At the centre of my proposed picture of reparation is the importance of the wrongdoer paying attention to, and recognising the individual significance of the harm done to the victim, and translating such an understanding into reparative action. Such actions can be expressive of genuine remorse, and provide the grounds for the victim to forgive the wrongdoer and thereby lift the burden of remorse.

3.8 Situating remorse in moral life

In the parody passage I quoted at the beginning of this chapter, Gaita suggests that, when we seek to understand what makes a principle or obligation *moral*, then we should seek part of the answer “in the kind of elaboration we give when we express most seriously our sense of what it means to wrong someone” (2004, xxi). Later, Gaita writes, “My argument is that remorse is a central and inextinguishable determinant of what it is for something to be a moral matter, and therefore of the sense of expressions like ‘morally wrong’, ‘morally ought’, ‘morally terrible’ and ‘morally the same as’” (2004, 60-61).

Consequently, remorse is not only central to our everyday understanding of what it is for something to be morally right or wrong, but should, thereby, be central to philosophical thinking about morality. And, philosophical thinking about morality, according to Gaita, should do at least two things: shed light on how a given account can deepen our thinking about morality, and “show that the account allows, to the reflective but the non-philosophical person, the possibility of an ever-deepening understanding of the nature of moral

significance" (2004, 41). Deeper understanding of moral significance cannot, therefore, be inconsistent with a philosophical account.

If remorse is an inextinguishable determinant of what it is for something to be a moral matter, then there is an important sense in which remorse, a philosophically neglected moral emotion, can illuminate philosophical thought about morality. Even if one does not agree with Gaita's most central claim, it is evident that his positioning of remorse, and his conceptualisation of it, afford the possibility of our learning more about the moral domain from a first-personal perspective, which is rather unusual in moral philosophy, where second- and third-personal perspectives typically dominate.

3.9 Conclusion

Remorse, according to Gaita, is the awakening to the harm of what one has done that focuses indivisible attention on the one who was wronged, and what, as wrongdoer, one has become. In this way, Gaita avoids the charge that the victim has disappeared from the picture. But, if we supplement his view with the idea that remorse is essentially a relational emotion, we can see that he also offers an important alternative to Taylor's excessively dualistic account. As I have emphasised, however, Gaita's account also contributes to the project of uncovering features of what remorse might be beyond the morality system. The contribution can be found in the key materials I think we ought to take from his account: the distinction between natural harms and moral significances of harms; corruptions or pathologies of genuine remorse; remorse as a form of moral understanding; and what these, taken together, make available for a distinctive conception of reparation. There are, also, questions regarding, for example, whether we want to accept a notion of blameless wrongdoing, what the scope of remorse is, and whether we want to accept the idea that remorse can be rationally independent of a sense of culpability. Some of these key materials and questions also appear, albeit in different forms, in Bernard Williams' work, which I will take up in the next chapter.

CHAPTER 4

AGENT-REGRET AND REMORSE BEYOND THE MORALITY SYSTEM

4.1 Introduction

[H]ow is one to chart the misunderstandings [of philosophers], without philosophical understanding of what the philosopher really meant?...[P]hilosophical insight is not something separate from the literary understanding of philosophical writing, because it is not separate from understanding philosophical writing at all...to identify what ideas are embodied in a text, particularly a philosopher's text, is no less a matter of philosophical comprehension than anything else.⁶¹

This passage from a review written by Bernard Williams describes what I am attempting to do in this chapter. Specifically, I chart the misunderstandings of philosophers regarding Williams' introduction and use of the term "agent-regret". To do so, I aim to capture what Williams really meant in his introduction of agent-regret by way of an exegesis of its first appearance in his paper "Moral luck" (1981b). What this close reading provides is not only a very different interpretation of agent-regret, but also the grounds for the introduction of an idea I take to be embodied in the text. My idea is that, contrary to the accepted interpretation of "agent-regret" as the term for the sentiment relevant in bad moral luck cases, we can understand it as a transitional, ultimately disposable concept, needed to get away from the morality system's way of understanding of first-personal expressions of moral responsibility. I go on to argue that, in light of this interpretation, we can realise the possibility of an expanded conception of remorse beyond the morality system. As we shall see, the argument grounds the idea for what I have been calling "the prospective genealogy of remorse".

Since its introduction in "Moral luck", the term "agent-regret" has consistently drawn the attention and interest of moral philosophers. What often structures this interest is the concern that Williams, in David Enoch's words, was "on to something deep, revealing, and important" (2012, 96). There are, of course, exceptions, and R.J. Wallace's recent *The View From Here* offers one such example. Though he explicitly takes inspiration from Williams specifically in

⁶¹ In Williams' review of *The English Moralists*, by Basil Willey (Norton, 1964) (2014, 53-54).

terms of his introduction of agent-regret, in light of his own theoretical interests, Wallace argues that “agent-regret” does not capture anything morally significant that cannot already be explained by his “more fundamental” category, “personal regret” (2013, 41). Though his inclusion of agent-regret is finally to argue for what he takes to be its irrelevance, in his characterisation of agent-regret, Wallace joins the company of other moral philosophers who mistakenly use what I call “the standard interpretation of agent-regret”. That the standard interpretation of agent-regret is a misreading of Williams not only prompts us to question those who, like Wallace, argue that the term is irrelevant; it also prompts us to question the use the notion is put to by those who embrace it.

Drawing on various accounts, I aim to show that the standard interpretation of agent-regret (henceforth, “SI”) is indeed a misreading, and one that is pervasive in the literature. In what follows, the formulation of the SI will be briefly outlined, followed by a critical analysis of Williams’ introduction of agent-regret in “Moral luck”, which ultimately reveals that the SI cannot be supported at the level of the text. In addition to revealing a widespread misinterpretation, the analysis allows for my claim that “agent-regret” is better understood as a transitional concept needed to get beyond the morality system in the territory of first-personal expressions of responsibility. It also makes available my minimal interpretive claim: that the philosophical structure “agent-regret” is not only a transitional platform for reflection on first-personal expressions of responsibility but also offers us, in its schematic character, the basis for a conception of remorse freed from the constraints of the morality system. In this way, this chapter serves to do what the opening quotation suggests: to better understand what Williams really meant by agent-regret in order to identify an idea, and a future-oriented genealogical possibility, I take to be embodied in his text(s).

4.2 The standard interpretation of agent-regret

The SI is fairly straightforward. In “Moral luck”, Williams presents the now famous example of the responsible lorry driver who accidentally hits and

kills a child. Williams argues that it is rational and appropriate to characterise the lorry driver's pained response to terrible harm he has wholly involuntarily caused as "agent-regret", and this example it is typically taken as not only illustrative but *paradigmatic* of what Williams takes agent-regret to be. For the majority of interpreters, "agent-regret" is nothing more than the term for a special sentiment or response that is rational and appropriate in cases of radically involuntary bad outcome luck. Some characteristic examples of the SI are as follows:

[T]he lorry driver of Williams's famous example acted in ways that brought about the death of a child, even though this wasn't something that he intended to cause, and even though he took all reasonable precautions in driving the vehicle to prevent an occurrence of this kind...**Agent-regret is meant to be the category of retrospective feeling that is peculiarly suited to one's involvement in unfortunate situations in cases of this kind** [my emphasis] (Wallace 2013, 34-35).

I'm afraid it is you. You are the driver from Williams's example. You hit a pedestrian, causing him serious harm. But the accident was not your fault....While we should all feel bad for the fate of the injured pedestrian, you, it seems, should feel that extra bit of agent-regret.... **we think that it is [in such a case] at the very least rationally and morally permissible to feel agent-regret** [my emphasis] (Enoch 2012, 97).⁶²

Though there is variability across accounts, those under the influence of the SI treat agent-regret as if it were best understood as (at least principally) the appropriate response in cases of wholly involuntary harm.⁶³ As I shall argue, this seriously distorts the role that Williams claims for it in ethical experience,

⁶² Enoch's characterisation of agent-regret not only associates it with the most involuntary cases but also as being an "extra bit" of feeling. The "extra bit" is relative to a spectator's response such that agent-regret is conceived of as something like "spectator's regret" plus an "extra bit of feeling". But, one of the most important points of Williams' discussion is precisely to emphasise the irrepressible first-personal perspective even in the most extreme cases of accidental or involuntary agency. It is, in other words, another error associated with the SI: to conceptualise agent-regret as essentially a spectator-like response with an extra bit of feeling.

⁶³ See Thomson (1993, 215n7), Gaita (2004, 53-54), de Wijze (2004, 461-463), and Jacobson (2013, 106-107; 112) for some other examples.

because his own presentation of agent-regret crucially also concerns cases in which moral bad luck strikes voluntary action.⁶⁴

Importantly, the SI does not, characteristically, stand alone. It is commonly interconnected with two other interpretive claims: one relating to what Williams has to say about remorse, and the other concerning what Williams has to say regarding a distinction between agent-regret and remorse. On the first claim, Williams is supposed to advocate the idea that remorse is, in accordance with the characteristic conception of it found in the morality system, rationally restricted to instances of voluntary agency. Though, in some cases, the claim is expressed in more implicit terms, others are made explicitly, such as the following:

Remorse is of course familiar to us as a peculiarly agential form of regret, appropriate especially to cases in which our actions attract moral objections of one kind or another. **But as Williams observed, its proper target is the class of performances that are voluntary** in some suitable sense (including both the things we have intentionally done, and the things that could have been foreseen to result from the things we intentionally did) [my emphasis] (Wallace 2013, 34)

Williams...thinks that remorse can rationally be directed only to voluntary actions (Gaita 2004, 53-54).⁶⁵

Conveniently, the claim that Williams upholds the prevailing idea that remorse is restricted to the voluntary works to support, and reinforce the SI. If Williams follows the widely uncontested claim that the scope of remorse is restricted to the voluntary, then he can be seen to be claiming that remorse is irrational and inappropriate in the most involuntary or accidental levels of agency like the lorry driver. It would seem to follow, in conformity with the SI, that the only moral sentiment that could be regarded as rational and appropriate in these unusual cases is “agent-regret”.

⁶⁴ There are important exceptions of course. See, for instance, Baron, who is entirely explicit that Williams claims that “agent-regret” can be felt in response to distinctive instances of voluntary agency: lesser of two evils cases. Baron takes there to be two types of agent-regret: (i) cases like the lorry driver where someone caused something bad to happen but not owing to any fault, and (ii) cases of lesser of two evils where the bad thing that one caused is more closely connected to a choice that one made (1988, 263).

⁶⁵ For more implicit examples see Wallace (2013, 41); de Wijze (2004, 461-463); Baron, (1988, 280n88) and Enoch (2012, 96).

This accounting of Williams encourages the second of the additional interpretative claims mentioned in the previous paragraph: the idea that Williams is advocating a sharp distinction between remorse and agent-regret. Namely, the SI and the remorse claim together, encourage the idea that remorse and agent-regret occupy parallel emotional domains sharply differentiated by the voluntary/involuntary (i.e., non-culpable involuntary, e.g., could not have foreseen what eventuated). On the one hand, according to this view, there is the characteristic conception of remorse as strictly rational and appropriate only with respect to the voluntary and, on the other hand, there is agent-regret, which captures only those special cases of non-culpable involuntary agency. For many interpreters accepting agent-regret under the SI, Williams' introduction of it presents a critical shift away from the distinction traditionally taken to adequately carve up the territory of first-personal expressions of moral responsibility, the one explored in Chapters 1 and 2, between remorse and regret.⁶⁶ As I constructed it in Chapter 1, the traditional distinction typically upholds the characteristic conception of remorse as the appropriate moral response in cases of voluntary breakings of moral obligations. This response, again, is contrasted with regret, which is understood as the rational and appropriate non-moral response in instances of involuntary agency where one has accidentally brought about harmful consequences or states of affairs.⁶⁷

Regardless of how agent-regret is accepted and theoretically accommodated, the acceptance of it, even under the SI, reflects philosophical acceptance that even in the most involuntary cases something *moral*, and more than mere non-moral regret, is rational and appropriate. In other words, when it is accommodated, agent-regret (even under the SI) reflects those philosophers' agreement with Williams that, even in such cases, we typically *do*

⁶⁶ This "new" distinction is widely seen as unseating the traditional distinction in the area of first-personal expression of responsibility for wrongdoing but, for some, the accommodation happens in the production of a picture, which organises the responses of remorse, agent-regret and regret into a spectrum (see de Wijze 2004; Baron 1988). Given Wallace's (2013) theoretical interest in a distinction between personal (remorse and agent-regret) and impersonal forms of regret, the spectrum or sliding scale model does not arise in his account.

⁶⁷ And, of course, as we saw in Chapter 1, non-moral regret is taken as rational and appropriate in tragic conflicts where one does something undesirable or bad in morally right action. But although Williams is not immediately concerned with the realism of moral conflicts in this connection, he does claim that agent-regret is appropriate in such cases of voluntary action (1981b, 31).

and *should* have a first-personal *moral* reaction to causing blameless harm.⁶⁸ The traditional relation between remorse and regret is seen as reconfigured by the introduction of this supposedly newly identified third response, agent-regret, because, as a first-personal response to involuntary actions and their consequences, it is neither remorse, nor regret. But more than this, as I will go on to argue, the accommodation of agent-regret under the SI also serves to preserve the conception of remorse that we previously found percolated in the morality system: MS-remorse. This is a conception of remorse that I will argue, contrary to broad philosophical opinion, Williams does *not* advocate.

4.3 The errors of the standard interpretation

4.3.1 *The forgotten voluntary cases of agent-regret*

Despite the ubiquity of the SI, a careful look back at Williams' writings quickly reveals its inaccuracy.⁶⁹ The first consideration that directly casts doubt on the idea that "agent-regret" refers primarily to radically involuntary cases is that Williams first introduces "agent-regret" not in the context of the lorry driver. Rather, he introduces the term in the context of an explicit interest in placing the kinds of thoughts that Anna Karenina—and others like her who ultimately suffer from a failure of luck intrinsic *to their chosen* project—"may be expected coherently to think about themselves" (1981b, 27). Anna Karenina, of course, makes the decision to leave her husband and her son in order to pursue a life with Vronsky. In pursuing her life with Vronsky, Anna remains conscious of "the cost exacted from others, above all from her son" for her decision (Williams 1981b, 26). Relative to her understanding at the time she makes her decision to leave Karenin and her son, we can suppose that, had things gone better in her life with Vronsky, she would have been better placed to live with

⁶⁸ Outlining the normative expectation of a first-personal expression of responsibility even in such no fault cases (as revealed by the lorry driver case) is the central preoccupation of Enoch's paper (2012). Enoch produces two explanatory options of how to accommodate Williams' insight: the moral luck explanation and his own, non-moral luck explanation. Enoch's paper, as I understand it, never reaches a point where he offers an argument for why the non-moral luck explanation has greater explanatory power than Williams' moral luck explanation. Namely, it just serves to present the reader with two options that turn on whether the reader accepts moral luck or not.

⁶⁹ The SI is not *everywhere*. For a closer reading of Williams on agent-regret see Raz (2012, 133-161).

this (painful) consciousness. But, as it turns out, her project with Vronsky fails and, significantly, it fails not because of some extrinsic misfortune such as Vronsky dying in an accident. Her project fails *intrinsically*, for the locus of luck is intrinsic to her project, lying significantly (if “by no means entirely”⁷⁰) in Anna herself—in “what is, and what is not, determined by her and by what she is” (Williams 1981b, 26).

The consideration that Anna’s project failed on these terms helps to mark the difference that Williams is concerned with outlining when he points to his interest in placing the kinds of thoughts Anna might be imagined to have at the intrinsic failure of her project. Williams writes,

I take it that as things are, her thought in killing herself is not just that [‘there is nothing more for me’] but relates inescapably also to the past and to what she has done. What she did, she now finds insupportable, because she could have been justified only by the life she hoped for, and those hopes were not just negated, but refuted, by what happened (1981b, 27).⁷¹

It is these kinds of retrospective thoughts about what one has done in instances of *voluntary agency* (e.g., Anna’s chosen project) that Williams is first interested in making plainer when he introduces “agent-regret” by way of distinguishing it from what he calls “regret in general”. When he writes “It is such thoughts that I want to place in a structure which will make their sense plainer” (Williams 1981b, 27), it is the retrospective thoughts of someone who, like Anna, has voluntarily done something that has failed owing to the risks intrinsic to the project. The psychological structure Williams introduces to make these retrospective thoughts plainer is what he calls “agent-regret”. As yet, Williams makes no mention of involuntary cases like the lorry driver.

Before embarking on placing the kinds of thoughts an agent might have at the intrinsic failure of their project into any particular structure, Williams also makes a frequently ignored methodological qualification: “The discussion is

⁷⁰ The qualification, “by no means entirely”, reflects the point that because her project is shared with Vronsky its success or failure lies partly outside herself, for it also depends on Vronsky and “what is, and is not, determined by him and by what he is” (Williams 1981b, 26).

⁷¹ Both luck extrinsic and intrinsic to a project are “necessary for success, and hence for actual justification, but only the latter relates to unjustification” (Williams 1981b, 26).

not in the first place directed to what we or others might say or think of these agents (though it has implications for that), but on what they can be expected coherently to think about themselves” (1981b, 27). As I interpret it, this qualification alerts one to Williams’ interest in creating some reflective distance between the first-personal and the second- and third-personal in this territory; something that is, as previously mentioned, unusual in such discussions (though as we saw in Chapters 2 and 3, a move Williams shares with both Taylor and Gaita).

For Williams, the aim is to temporarily suspend the ever immediate question of how a victim, a third party or even a reader of the example might (imaginatively) respond to such an agent in such circumstances. This is in order to consider what can be made of the kinds of thoughts and feelings such an agent might reasonably be thought to have from their own (first-personal) perspective. The difference is subtle, but I take it that Williams is interested here, first, in the question of a description from the first-personal perspective, and not immediately the persistently proximate question of what we or imagined others might say or think of these agents (and normatively speaking, what we think agents, victims and third parties should say, think or feel). The point of doing this is, as cited above, to create a “structure” which, in line with Williams’ explicit philosophical ambitions more generally, can help us to better understand our thoughts and feelings in situations like these.

4.3.2 The scope of agent-regret

Keeping these methodological considerations in mind, we can see how Williams’ use of “agent-regret” to first make sense of voluntary cases is further supported by his claim regarding the scope of agent-regret. Immediately following the distinction between agent-regret and regret in general, and, significantly, *before* introducing the lorry driver example, Williams writes:

The sentiment of agent-regret is by no means restricted to *voluntary* agency. It can extend far beyond what one intentionally did to almost anything for which one was causally responsible in virtue of something one intentionally did (1981b, 28).

Along with the fact that the notion of agent-regret first grows from the consideration of cases of voluntary agency, the first line of this description of the scope of agent-regret straightforwardly refutes the SI as referring exclusively, and/or paradigmatically, to the most involuntary of cases. His clear statement that the sentiment of agent-regret “is by no means restricted to *voluntary agency*” [Williams’ emphasis] reflects the fact that his exploration of agent-regret starts with cases of moral luck operating in consort with voluntary action. The extension of agent-regret from the voluntary to the most involuntary of cases is then explicated in the second line of the quotation, which describes the scope of agent-regret as extending “far beyond” what one has intentionally done to “almost anything for which one is causally responsible in virtue of something one intentionally did”. Aside from straightforwardly refuting the SI idea that the scope of agent-regret is restricted to, or even primarily associated with, cases of involuntary agency, the breadth of the scope of agent-regret also suggests that it is doubtful that Williams even thought there was such a thing as a paradigmatic example of agent-regret, let alone the case of the lorry driver.

It is worth making some further remarks on what Williams can be read to be claiming regarding the scope of agent-regret. A prevalent view of moral responsibility takes it as a necessary condition that, in any particular case, the agent could reasonably be expected to have known or foreseen that the consequences or state of affairs would have followed from their action. In their respective accounts, Wallace and Baron are two such theorists. But Williams is pushing the epistemic boundary to its limit. The agent need not be in a position to foresee anything: the mere retrospective acknowledgement “I caused this harm” is sufficient for appropriate agent-regret, and so for the moral responsibility of which it is an expression (recall Gaita’s broad conception of moral responsibility). With this detachment from any condition of reasonable foreknowledge established, it is clear why sentiments of agent-regret extend naturally to even radically involuntary levels of agency like the lorry driver’s case. While, in the abstract, the lorry driver will have been in a position to know that an activity like driving carries such hazards, still there are no grounds at all to regard him as ought-ing to have known that, in this particular instance, it would be instantiated. He could not possibly have known that the child would

run into the road at that moment, and yet, on Williams' picture, he is morally responsible for the terrible thing he caused, and while blame is out of order, agent-regret is appropriate.⁷²

Following this statement of the broad scope of agent-regret, and immediately before describing the lorry driver case, Williams writes, "Yet even at deeply accidental or non-voluntary levels of agency, sentiments 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"(1981b, 28). As this analysis has so far shown, in its introduction, Williams' initial concern when he differentiates agent-regret from regret in general is with respect to cases of voluntary agency. Now, here, to support his claim regarding the scope of agent-regret as extending far beyond what one has intentionally done but delimited by the condition of causal responsibility, Williams returns to the contrast with regret in general: even at the very limit of involuntary agency, sentiments of agent-regret are still importantly distinguishable from regret in general. Williams introduces the lorry driver case to illustrate this point. He explains that the driver is going to feel differently from any spectator in that the driver will have an agent's thought; that perhaps *he* could have prevented the death of the child, something not available to the spectator. This is, in the first instance, what differentiates agent-regret (in voluntary or involuntary cases) from regret in general: "the supposed possible difference is that one might have acted otherwise, and the focus of the regret is on that possibility, the thought being

⁷² For Williams, all intentional action is subject to luck in its success or failure, some of it internal, some of it external. "The examples of Gauguin and Anna Karenina are, of course, cases of voluntary agency, but they share something with the involuntary cases just mentioned, for the 'luck' of the agents relates to those elements which are essential to the outcome but lie outside their control, and what we are discussing is in this way a very drastic example of determination by the actual, the determination of the agents' judgments on their decisions by what, beyond their will, actually occurs"(1981b, 30).

formed in part by first-personal conceptions of how one might have acted otherwise” (Williams 1981b, 27).⁷³

In terms of the lorry driver, Williams describes the first-personal expression of how he might have acted otherwise in the following way, “He deeply wishes that he had made that change which, had he known it, was in his power and which would have altered the outcome” (1981, 30). This is a clear expression of the epistemic limit of the scope of agent-regret outlined above: he need not have been in a position to have known the child would dash out onto the road in order for him to feel agent-regret. This makes plain an important point against the SI: the lorry driver case, as an example of non-culpable involuntary action, does not arise as the primary example of agent-regret but, rather, as an illustration of how far the scope of agent-regret extends. The lorry driver example is not, therefore, the primary or paradigmatic example of agent-regret, but is better understood as a *limiting* case of agent-regret.⁷⁴

The status of the lorry driver example as the limiting case in Williams’ account is crucial with respect to the SI. Up to this point, I have shown that the SI errs in forgetting Williams’ remarks about the voluntary with respect to agent-regret but also, and often relatedly, errs in taking the lorry driver example to be the primary and definitive case of agent-regret rather than the limiting case. Even if an interpreter does not forget Williams’ remarks about the voluntary, but nevertheless takes the lorry driver case to be paradigmatic of agent-regret, this still encourages a misconception of the domain of agent-regret

⁷³ Agent-regret is importantly not just distinguished from regret in general along this central line of first-personal subject matter. It is also distinguished by having “a particular kind of psychological content” as well as “a particular kind of expression”. The particular kind of expression of agent-regret relates to the character of the desire to make reparations, which Williams explains through his insurance test (1981b, 28-29). The insurance test asks whether, from the agent’s point of view, insurance money will satisfy with respect to compensation or reparation. An agent who views the loss or harm as being the sort of thing that cannot be adequately compensated for, if at all, by an insurance pay-out, is likely to continue to feel like she should do something “because her actions might have some reparative significance other than compensation” (1981b, 29). To get this idea through, Williams is relying on the difference between compensating for the natural effects of the harm or loss, and the significances of harms (Gaita 2004, 54).

⁷⁴ Strictly speaking it is not quite *the* limiting case of agent-regret. This is because Williams allows for a less direct or even indirect causal connection between the outcome and one’s action by affording the possibility that, within an unstated limit, being a participant is sufficient for agent-regret (1981b, 27). This qualification, I take it, allows for an extension of the intelligible scope of causal responsibility from direct causal links to cases of collective action, and cases like Gaita’s Dutch woman, where one’s action is a significant link in a causal chain that lead to the harmful thing.

by focusing excessively on the involuntary. But, as the passage above regarding the broad scope of agent-regret and the introduction of agent-regret in terms of voluntary cases show, the paradigmatic case of agent-regret is *not* the involuntary case. If anything, these remarks taken together suggest that, if there were such a thing as a paradigmatic or definitive case of agent-regret, it would more likely be an instance of voluntary agency, such as Anna Karenina.

It is no accident that the radically involuntary case of the lorry driver comes after the example of Anna. For, in echoing Williams' methodological staving off the impulse to make second- and third-personal moral judgements, the lorry driver case is such that it ensures that even the most avid fault-finders will have to acknowledge they cannot find fault. There is a clear difficulty in devising examples as unambiguous as the lorry drivers, i.e., where no one would reasonably think that the agent really could have known or foreseen, in the particular instance, that their actions would result in such terrible consequences. Despite this difficulty, the example offers an attractive analytic simplicity that the voluntary cases like Anna's cannot. This appealing analytic simplicity may go some way in explaining the SI's focus on the lorry driver example as paradigmatic of agent-regret; not only is explaining cases like Anna's more difficult in light of its potentially diverting invitations to moral judgement but, perhaps equally, it can be difficult to conceptualise basic connections between voluntary and involuntary cases.

The lorry driver example serves not just to illustrate the outer limit of agent-regret but, also, to remind us that, by appeal to actual experience through detailed description, it is (presented as) a *fact* that we experience agent-regret in even the most involuntary of cases (something Enoch and others accept). Addressing how Williams takes "the fact of agent-regret about the involuntary" to affect the traditional dichotomy between remorse and regret has the upshot of directly undermining one of the two interpretive claims associated with the SI: that Williams claims remorse is restricted to the voluntary. Again, this claim is not supported at the level of the text. There is only one place in "Moral luck" where Williams explicitly refers to remorse, and it is in the following passage:

Agent-regret about the involuntary also helps us to get away from a dichotomy which is often relied on in these matters, expressed in such terms as *regret* and *remorse*, where “regret” is identified in effect as the regret of the spectator, **while “remorse” is what we have called “agent-regret”, but under the restriction that it applies only to the voluntary.** The fact that we have agent-regret about the involuntary, and would not readily recognize a life without it (though we may think we might), shows already that **there is something wrong with this dichotomy: such regret is neither mere spectator’s regret, nor (by this definition) remorse** [my emphasis] (1981b, 30).

Is Williams really claiming that remorse should be understood as restricted to the voluntary? Contrary to SI, a careful reading shows that he is not making this claim at all. The first point to note is that Williams is interested in *getting away from* a dichotomy “often relied on in these matters, expressed in such terms as between *regret* and *remorse*”.

Utilising his own terminology, Williams describes the regret of the dichotomy as what he has called “the regret of the spectator”, and the remorse of the dichotomy as what he has called “agent-regret under the restriction that it applies only to the voluntary”. This matters because it signals that the conception of remorse operating is that found in the dichotomy, and not Williams’ own conception of remorse. From this terminological difference, the fact that he wants to move away from the dichotomy, and that he has established his interest in arguing against the idea that what he has called “agent-regret” should be assumed to be restricted to the voluntary, we can conclude that he is *not* endorsing the conception of remorse which sees it as restricted to the voluntary. In fact, the conclusion we should draw here is that Williams wants to move away from a narrow conception of remorse that restricts it to the voluntary. This conclusion raises an important question: how did we become burdened with the misguided dichotomy that Williams diagnoses? The answer I propose is that the narrow conception of remorse, and the dichotomy that gives rise to it are, for Williams, artefacts of the morality system.

4.4 Undoing the morality system's dichotomy

The claims I am putting forward are bolstered by Williams' more general critical interest in questioning and eroding the centrality of the idea of the voluntary in moral philosophy. More specifically, the idea of the voluntary as playing a cardinal role in structuring our moral psychology he finds problematically advocated by the morality system. In a footnote on his chapter on the morality system in *ELP* Williams writes that, "It [his paper "Moral luck"] illustrates the general point that the morality system lays particularly heavy weight on the unsure structure of voluntariness" (2006a, 243). When he makes mention of this it is in the context of discussing the dichotomy between regret and remorse advocated by the morality system (henceforth "MS-dichotomy"⁷⁵). As we saw in Chapter 1, on the morality system's terms, remorse is the purely moral rational reaction to voluntary breakings of moral obligations, whereas the regret of the MS-dichotomy is the appropriate response to unfortunate states of affairs one brings about involuntarily or through no fault of one's own. Once again, it is, on this view, understandable that the lorry driver feels badly for what he has done, but this cannot be understood as a bad *moral* feeling for he has not *voluntarily* broken a moral obligation (Williams 2006a, 177). Again, the lorry driver has done nothing morally blameworthy, and so any response he ought to have to what he has brought about must be non-moral in kind. The bad feeling is then characterised as non-moral regret, which can, on Williams', and Taylor's analysis of it, be felt by anyone who has knowledge of it.

The non-moral regret of the morality system encourages the idea that for an agent to be rational, morally speaking, she ought to treat what she caused or was involved in bringing about as a mere spectator would. In other words, as we saw in Chapter 1, in cases of non-culpable involuntary agency, the morality system advocates as rationally and morally correct an impersonal non-moral response to something terrible or awful I have done. Unless the action was voluntary, any action that causes harm or loss to others cannot be morally significant to the agent first-personally. As Williams puts the point in *ELP*, "The

⁷⁵ In Chapter 1, I referred to the *distinction* between regret and remorse found in the morality system. Here, I will follow Williams and use "dichotomy", although I take the intended contrast between regret and remorse to be the same.

thought *I did it* has no special significance; what is significant [in the system] is whether I voluntarily did what I ought to have done” (2006a, 177). This leads Williams to claim that: “This turns our attention away from an important dimension of ethical experience, which lies in the distinction simply between what one has done and what one has not done. That can be as important as the distinction between the voluntary and the non-voluntary” (2006a, 177). The moral psychological distortion this generates is evidenced by the consideration that even in the most extreme cases of accidental agency the response of agent-regret, and not the regret of the spectator, is typical and, in some way, expected.⁷⁶

Here, we return to an important point regarding the errors of the SI. Williams is using *the fact* that we do readily and familiarly experience some emotion over instances of involuntary agency that is not impersonal, and not non-moral. This is, in part, to shift the regret side of the MS-dichotomy: to move away from the idea that non-moral regret is the rational and morally appropriate response in the most accidental or involuntary of cases. That is, whatever emotion the lorry driver is experiencing, and that we expect him to feel in response to what he has brought about is not non-moral regret. This does not mean, of course, that Williams is saying it is not remorse either. It follows, therefore, that if Williams’ appeal to “actual experience” in the lorry driver case convinces one that, in practice, we do recognise a significant difference between the regret a spectator might feel, and the regret an agent will likely feel, then, accepting this as a fact can successfully unseat the role of regret in the MS-dichotomy. But, its significance is not limited to this claim, and this is often missed, for it also puts pressure on the idea that remorse is, in these cases, out of order.

Recall the second interconnected interpretive claim mentioned in the first section; that the SI characteristically encourages the idea that Williams is advocating a distinction between remorse and agent-regret that sees them as

⁷⁶ This is central to Enoch’s (2012) concern. Following Williams, Enoch accepts that, in practice, even in accidental cases there is an as yet undistinguished ethically tinted expectation that structures third-personal responses to the harm the agent accidentally brought about. It is his central concern in the paper to differentiate the character of such an expectation and the normative standards associated with it. More specifically, perhaps, is the idea that we expect others to take responsibility even for things that we would not blame them for.

sharply distinguished by the voluntary/involuntary. Now, we are in a position to diagnose why the SI encourages this idea of a “new” distinction between remorse and agent-regret. In the lengthy passage in the previous section, Williams claims that “the fact that we have agent-regret about the involuntary...shows that there is something wrong with this [remorse and regret] dichotomy”. Given that the upholders of the SI tend to maintain the idea that agent-regret is paradigmatically felt in the most involuntary of cases *and* that remorse is restricted to the voluntary, they only see agent-regret as showing the regret side of the MS-dichotomy to be wrong. That is, under the SI, the remorse side of the MS-dichotomy is taken as fixed, and the regret side as moveable not in light of *the fact* that we experience agent-regret about the involuntary (as Williams claims), but *because* agent-regret under the SI is taken to be the appropriate term for (at least paradigmatically) the non-culpable involuntary cases. Keeping this in mind, it then seems clear how, under the SI, when Williams claims that the fact that we have agent-regret about the involuntary helps us to get away from the MS-dichotomy, the “getting away from the dichotomy” is read by SI supporters as establishing a new distinction between remorse and agent-regret.

Another way of expressing this is to say that Williams is interpreted as claiming that agent-regret (under the SI) shows that something is wrong with the MS-dichotomy only on the regret side of it, and that replacing the regret side with “agent-regret” counts as moving away from the MS-dichotomy. This, I take it, explains the idea of a shift to a new distinction. Aside from the fact that, contrary to Williams, the new distinction preserves the centrality of the structure of the voluntary, this elaboration shows up a key error. If we understand agent-regret as Williams clearly intended it, as referring to the voluntary as well as the involuntary, then Williams’ use of *the fact* that we experience agent-regret about the involuntary takes on a very different significance with respect to what it shows to be wrong about the MS-dichotomy, and how it helps us to move away from it. How to interpret that significance is what I will consider in the rest of this chapter.

The idea that agent-regret about the involuntary unseats the regret side of the MS-dichotomy is not incorrect. But, this is certainly not all the fact of

agent-regret about the involuntary is supposed to do with respect to the MS-dichotomy. When Williams writes about the fact of the involuntary showing that there is something wrong with the MS-dichotomy between remorse and regret, he makes another important qualification: “such regret is neither mere spectator’s regret, nor (by this definition) remorse”. The qualification “by this definition” brings us back to what was explicated above: the conception of remorse as restricted to the voluntary or, in Williams’ terms, “agent-regret but under the restriction that it applies only to the voluntary” (1981b, 30). The qualification signals, again, that this is not Williams’ conception of remorse, and that the fact of agent-regret about the involuntary should also make us question the other side of the MS-dichotomy: its conception of remorse as restricted to the voluntary. But, I think this is not all the qualification suggests. The qualification also tantalisingly indicates the possibility that there could be a form of remorse beyond the constraints placed on it by the morality system. This possibility is what my idea of a prospective genealogy picks out: a form of remorse (as both moral experience and moral concept) that is released from the bind to the voluntary that is imposed by the morality system.

4.5 What is the role of agent-regret in Williams?

In his own terms, Williams characterises the MS-dichotomy as being between a form of regret in which one takes the view of a spectator with respect to one’s past actions, and remorse as what he has described as agent-regret but under the restriction that it applies only to the voluntary. If we understand Williams as taking the scope of agent-regret to extend from the voluntary to include almost anything for which one was causally responsible in virtue of something one intentionally did, then the fact that we experience agent-regret about the involuntary should not be read as picking out some third relevant sentiment that disrupts the MS-dichotomy. Rather, the fact that we experience agent-regret about the involuntary not only shows there is something wrong with the regret side of the MS-dichotomy, but, equally, it shows that there is something wrong with the MS-dichotomy insofar as it supports a definition of remorse that sees it as restricted to the voluntary.

If Williams is claiming that the scope of agent-regret extends far beyond the voluntary, and remorse in the MS-dichotomy just is agent-regret restricted to the voluntary, then the claim can be read in the other direction: the fact that we have agent-regret about the involuntary, as well as the voluntary, suggests that there is something wrong with the picture of remorse in the morality system that restricts it to instances of voluntary agency. Drawing out this implication is what drives my novel interpretation of the role of “agent-regret” in Williams’ work. As I will go on to claim, I take “agent-regret” to emerge as a transitional concept that can lead the way to an expanded conception of remorse beyond the morality system’s constraints.

Leaving aside the difficulties of Williams’ writing, and the ways it can be read to encourage the SI and the interconnected interpretive claims, let me offer a partial diagnosis of how the SI has arisen and acquired some of its significant authority. The misreading arises, at least in part, owing to a moral philosophers’ habit of reading back into the text the morality system’s conception of remorse as restricted to the voluntary. The unquestioned belief that remorse is restricted to the voluntary goes some way in explaining the origins and persistence of the SI’s construal of agent-regret as (at least paradigmatically) applied to instances of involuntary agency. Moreover, it likely goes some way in explaining why the SI has gained acceptance as capturing a distinctive phenomenon for, read as restricted to or paradigmatic of the most involuntary cases, it is prevented from being seen as undermining MS-remorse. In fact, when agent-regret is understood in terms of the SI it can implicitly reinforce the morality system’s conception of remorse as restricted to the voluntary. In this way, we can see that, though the SI can be seen to reflect philosophers’ agreement with Williams that even in the most involuntary cases we typically do, and should, have a first-personal *moral* reaction to causing blameless harm, the SI blocks the possibility of understanding the full critical nature of Williams’ use of agent-regret as equally extending to MS-remorse.

One might say that the force of the SI, the restriction of remorse to the voluntary, and the effect it has on distorting Williams’ account of agent-regret, are all things which reflect the intractable character of the other idea Williams is aiming to question: the central role of a particular conception of the voluntary in

organising our moral responses (particularly with respect to first-personal expressions of moral responsibility). It is important to note the character of Williams' critique of the voluntary as he finds it in the morality system. As I presented it in Chapter 1, Williams argues against the morality system's conception of the purely or deeply voluntary but, importantly, he also argues against the morality system's centralising of the voluntary in structuring our moral psychology. As I am interpreting Williams, his introduction of "agent-regret" needs to be understood as a key part of his broader critique of how key elements of our moral psychology are constructed, constrained and "purified" in the morality system.

Though, throughout his work, Williams offers an alternative "superficial" conception of the voluntary, he nevertheless maintains his argument for the diminished importance, though by no means irrelevance, of the voluntary in organising our moral psychology.⁷⁷ That these critical points have been missed is evident in the shift the SI encourages to a revised distinction, for it preserves the distinction between the voluntary and involuntary that is familiar from the MS-dichotomy. But, as my analysis showed, Williams does not make the claim himself, and nor does he understand agent-regret to refer exclusively to involuntary cases. That this is the case suggests, rather, that the distinction between the voluntary and involuntary is a red herring with respect to understanding any distinction Williams may or may not have intended between remorse and agent-regret. For when Williams says he wants to move away from the dichotomy he means that he wants to get away from a dichotomy in this territory altogether; not simply refocus it or shift it, as the SI would have him do. This makes available the move I am urging here: namely that we understand agent-regret as indicating the possibility of moral psychological formations beyond the morality system, and its dichotomous conception of remorse and regret.

⁷⁷ One version of Williams' notion of the voluntary is the following: "Out of these materials it is possible to construct a notion—an inherently vague and limited notion—of the voluntary: a certain thing is done voluntarily if (very roughly) it is an intentional aspect of an action done in a normal state of mind" (1993a, 66). For others, which are basically the same, see Williams (2006a, 194; 1995b, 25; 1995e, 242). For a recent, though ultimately mistaken, critique of Williams on the voluntary, see Hyman (2013).

Let me try to bolster this perspective on agent-regret a little more. Agent-regret is clearly a far broader category than the SI would have us think, but is there anything specific in Williams' texts that encourages the idea that it indicates a future beyond the morality system? Some suggestive remarks can be found in "Moral luck: a postscript":

We feel the need to exempt agents from (some) blame for (some) things done involuntarily and also, perhaps, to deepen the idea of the involuntary, because we think it unjust not to do so...It can readily be pointed out that this idea, though it certainly expresses morality's concerns, can hardly be deployed to back them up: the notion of injustice at work here is morality's own. That is correct, but it is not very helpful. If these are our feelings and our dispositions of judgement, then that is what they are, and a historical or philosophical distinction between morality and a more general conception of the ethical is not, in itself, going to make them go away. We need, not the general formula, but insight into what the distinctions of the voluntary and the involuntary, and the other conceptions related to the avoidance of luck, mean to us. Among the questions raised here are the following. Since, necessarily, we cannot ultimately avoid luck, what do we actually do about it? Why do we mind more about it in some connections than in others? Among our reactions to things that are done wrongly or badly, what does blame, in particular, do? Does it apply in the same way to others and to one's own self? (Williams 1995e, 243).

The first aspect of this passage to note is the clear interest Williams expresses in moving away from the morality system particularly with respect to "the general formula". The general formula or the doctrine of the voluntary is, again, yet another expression of the morality system's insistence on the just application of blame, i.e., blame is only justified when applied to instances of voluntary agency (where this is the purely voluntary conception of the morality system). In this way, as we have seen, in the morality system, moral responsibility is conceived of as coextensive with the scope of possible/intelligible/fair blame. When Williams claims that "we do not need a general formula but insight into what the distinctions of the voluntary and involuntary mean to us", and poses the final questions there is the implication that he is trying to provoke reflection. The prompt to reflection, he suggests, may go some way in helping us to better

understand the operations of luck, and many of our moral concepts in relation to it.⁷⁸

Taking the interpretation of agent-regret I have set out above, together with this passage, and others like it peppered across Williams' work, I want to claim that "agent-regret" is best understood as a term of art in Williams' wider project of moving away from the distortions of the morality system. Specifically, I propose we come to understand the term "agent-regret" as a *transitional concept* that Williams creates that facilitates the disruption of the distorting strictures of the morality system, and affords the possibility of reflection on alternative possibilities. In this way, it can also help us to consider what the distinction between the voluntary and involuntary means to us at the level of the first-personal, and how, in this connection, it might relate to the other distinction Williams takes to be of equal importance: that between what I have done and what I have not done (2006a, 177).⁷⁹

If agent-regret is a transitional concept needed to move away from the morality system's excessive abstraction and idealisation of first-personal expressions of moral responsibility, then it makes sense to begin, as I did in the first chapter, by identifying the morality system's conception of remorse. It follows, on the interpretation I have been elaborating here, that we can use the structure of agent-regret to step back from MS-remorse, question it, and reflect on the conceptual space of first-personal expression of moral responsibility. In doing so, and looking to other conceptions like the one offered by Raimond Gaita, we can imagine what an expanded form of remorse freed from the constraints and demands of the system might be like. It is in this sense that my interpretive project here amounts to a *prospective* genealogy of remorse.

⁷⁸ One dimension, particularly relevant to the immediate discussion of remorse is Williams' final question: whether blame and self-blame/remorse are symmetrical in the manner proposed in Chapter 1. For an interesting discussion of the possibilities of an asymmetry between blame and self-blame see Moran (2001).

⁷⁹ Williams offers some further advice for reflection in this territory when he writes: "[I]f we want to understand why it might be important for us to distinguish the harms we do voluntarily from those that we do involuntarily, we shall hope to succeed only if we ask what kinds of failings or inadequacy are the source of the harms, and what those failings mean in the context of our own and other people's lives. This is the territory of shame; it is only by moving into it that we may gain some insight into one of the main preoccupations of the morality that centres itself on guilt" (1993a, 94).

4.6 Remorse by any other name?

There are two further interpretative claims I want to suggestively leave open for future exploration: the minimal interpretive claim, and the strong interpretive claim. The minimal interpretive claim is simply the idea that the philosophical structure “agent-regret” is not only a transitional platform for reflection, but may also offer us materials for the basis of a conception of remorse beyond the morality system. This would at first involve a deeper examination of Williams’ remarks on agent-regret, which would involve exploring guilt, for Williams claims that agent-regret “can be psychologically and structurally a manifestation of guilt” (1993a, 93). A helpful passage in this connection that exemplifies the possibilities is the following,

This is clearly illustrated by a point I have just mentioned, that guilt, insofar as it concerns itself with victims, is not necessarily (93) or obviously restricted to voluntary actions. I may rightly feel that victims have a claim on me and that their anger and suffering looks towards me, even though I have acted involuntarily. The conceptions of modern morality, however, insist at once on the primacy of guilt, its significance in turning us towards victims, and its rational restriction to the voluntary. It is under considerable strain insisting on all these things at once. In fact, if we want to understand why it might be important for us to distinguish the harms we do voluntarily from those that we do involuntarily, we shall hope to succeed only if we ask what kinds of failings or inadequacy are the source of the harms, and what those failings mean in the context of our own and other people’s lives. This is the territory of shame; it is only by moving into it that we may gain some insight into one of the main preoccupations of the morality that centres itself on guilt (Williams 1993a, 94).

There is a great deal packed into this short passage but, for my present purposes, it simply offers an example of how to re-think guilt and remorse in light of the disconnection Williams promotes between harm and voluntary action. Williams clearly offers a way forward with respect to how to think about the voluntary and the involuntary with respect to guilt, and remorse—by looking to shame (1993a, 75-102; 219-222). Other related avenues of examination in this connection are Williams’ model of responsibility (1993a, 50-74; 2006d, 66-68; 2006b, 119-125), the voluntary (1993a, 66-70; 1995b; 2006b, 119-125), and blame (1995c).

In the examination of these other materials one may uncover support for the strong interpretive claim, which takes seriously the possibility that what Williams calls “agent-regret” can be understood as what would otherwise be remorse in ordinary experience. Namely, by looking at what Williams has to say about agent-regret, we may discover that its philosophical structure is what would be remorse were it not for the philosophical impositions and distortions generated by moral philosophers in replicating the morality system. Admittedly, in the immediate context of “Moral luck”, the plausibility of this strong interpretive claim relies heavily on Williams’ claim that MS-remorse is “what we have called ‘agent-regret’, but under the restriction that it applies only to the voluntary” (1981b, 30). However, to test this interpretive hypothesis would require even more work than the minimal interpretive claim, and yet still involve a larger interpretive leap. Furthermore, even if one were to elaborate all this material a very likely conclusion would be that “agent-regret” is simply too thin a structure to capture what we want of a broader account of remorse. Perhaps, what agent-regret does best is to draw our attention to the fact that remorse, as a first-personal form of moral responsibility, can be appropriately experienced simply in virtue of what I have done or not done (Williams 2006a, 177). Where to go from there is offered by the possibilities made available in Williams’ other remarks.

It is important to note that whatever results from this exploration, so long as one is committed to trying to capture what Williams is saying, the results will not be absolutely definitive in character. As one can already see with agent-regret, it is a broad, capacious concept that has more the character of a psychological field or landscape than a clearly defined conceptual shape (Williams 1993a, 89). As Christopher Cordner remarks in relation to Williams’ account of guilt, “Williams does not confine himself to conceptual analysis in his account of guilt. This is at least partly because he does not think there is a single concept of guilt. Historically, there have been different conceptions of guilt, aligned with correspondingly different experiences of it” (2007, 340). For, in the realm of moral psychology where Williams is concerned to draw from immediate experience, its thought and feelings, Williams is never engaged in a project of determining definitive conceptions of guilt or shame or agent-regret

(i.e., *the* conception of guilt or *the* conception of shame). Any results of an attempt to construct Williams-esque models of these moral emotions will therefore, not have the absolutely definitive character Gabriele Taylor, for example, is after.

In the introduction to this thesis, I claimed that what makes remorse a *moral* emotion is the distinctive relation it bears to those harmed by one's actions and their consequences. As we saw, this claim finds expression in Rawls' "association guilt", and in Taylor's and Gaita's conceptions of remorse. This claim also helped to reveal a key insight: that remorse is relational. Namely, remorse exhibits at once an outward and inward trajectory because it relates the subject and object of a harmful or wrongful action. Moreover, this claim served to anchor the central objection against MS-remorse, as well as Taylor's conception of guilt.

As I argued across the chapters, if we accept and *prioritise* the idea that remorse has a special connection with those harmed, then this raises doubts about the plausibility of the claim that remorse is restricted to the voluntary. In one sense, this is what I take Williams to have been attempting to show by making "agent-regret" such a capacious category—the very fact that one has harmed another, as in the lorry driver case, may be sufficient for intelligible, coherent and reasonable remorse. This is an expression of what Williams says is "the utterly familiar fact that what has happened to others through our agency can have its own authority over our feelings, though we brought it about involuntarily" (1993, 92). Given the incitement to philosophical inquiry I take this to present, I want to claim that our further explorations for a conception or form of remorse, freed from the distortions of the morality system, ought to begin with the idea that *remorse is a first-personal expression of responsibility for harm done to others*.

Importantly, this formulation of remorse is neutral with respect to the voluntariness or intentionality of the harmful action and its consequences. That is, we can consider how or whether or to what extent the voluntary can be built-in or adjusted or explained in relation to this formulation (which, depending on our answers, may dissolve the formulation). Presently, there are at least two currently identifiable routes to take with respect to this issue. In one direction,

we could follow many philosophers who now accept a broader notion of the voluntary as extending to culpable ignorance and take this notion to restrict remorse or guilt. Alternatively, we could accept the much broader scope of remorse or guilt as extending to non-culpable ignorance that both Williams and Gaita advocate. This broader scope is focused on the distinction between what I have done and what I have not done, and not strictly on the distinction between the voluntary and involuntary. Neither philosopher is claiming that the voluntary or the intentional is unimportant in this connection. However, both Williams and Gaita claim that when one feels remorse or guilt for a harmful action that was voluntary or intentional, in terms of moral psychology, the appropriate place to look for the importance of the voluntary is *shame*.

Through the notion of a prospective genealogy, I have questioned an existing stronghold conception of remorse, and developed grounds for a broader conception of remorse than is currently standardly available in moral philosophy. In an important sense, I have raised more questions than I have provided answers. In Chapters 1 and 4, for example, we encountered the question of whether remorse is best understood as the first-personal mirror of blame or as bearing an asymmetrical relation with blame. Relatedly, in Chapter 2 we encountered the question of whether remorse has a necessary connection with culpability. In Chapter 3, we encountered the questions of what the relations between harms and wrongdoing amount to, and whether intelligible remorse conditions what we can see as wrongdoing. I take it that raising questions, as I have done in this thesis, follows from an attempt to foreground an emotion that has so far been philosophically neglected. For, surely, part of what it is to argue for the need for philosophical inquiry is first to uncover what questions and puzzles there are to ask or to consider. In this way, even the questions raised in the thesis do not leave us bereft of materials to work with.

I have shown that moral philosophers with an interest in moral psychology ought to consider more closely what is at stake in the rich and underexplored territory of first-personal expressions of (moral) responsibility. And, although it is too early to say how the prospective genealogy of remorse would end, on my reading of agent-regret and the various constructive materials offered throughout this thesis, we can see how it might end. Namely,

we can see how remorse has been pressurized into a highly confined form, and how, once those pressures are lifted, we can discover a broader, arguably less moralistic, form of remorse.

4.7 Conclusion

What is not in doubt is that Williams is interested in getting away from the MS-dichotomy between remorse and regret. What is in doubt is how Williams' term "agent-regret" is typically interpreted, and how that understanding is said to relate to the MS-dichotomy. I have suggested that a closer reading of the text shows two prevalent interpretive ideas to be quite false: the idea that "agent-regret" is meant to apply primarily to cases of radically involuntary agency; and the idea, in the other direction, that Williams takes remorse to be restricted to the voluntary. If the scope of the sentiment of agent-regret is understood as inclusive of cases of voluntary and involuntary agency alike, then the MS-dichotomy can be questioned in both of these directions. In terms of the non-moral (spectator-like) regret side, the fact that we experience agent-regret over the most involuntary cases casts doubt on the ideas upheld by the morality system that, in both involuntary and lesser of two evils cases the appropriate (and rational) response to what we have done is both (i) non-moral in character despite the fact that we have harmed another in what we have brought about, and (ii) that it can be of no special significance with respect to what I have done that I am the one who did it. Moreover, if agent-regret has as broad a scope as I have proposed, then the definition of remorse found in the MS-dichotomy (as agent-regret restricted to the voluntary) is opened to questioning. Namely, if agent-regret is felt regarding voluntary and involuntary actions alike, and MS-remorse just is "agent-regret only restricted to the voluntary," then we can propose that MS-remorse is falsely restricted to the voluntary.

If we accept "agent-regret" as a concept introduced by Williams in an attempt to convince moral philosophers of the morality system's distortion in the field of first-personal expressions of moral responsibility, there is a further conclusion waiting to be drawn. The conclusion concerns the form that remorse

might take once liberated from the morality system: there is a possible understanding and experience of remorse that would be entirely appropriate and rational in radically involuntary cases, such as that of the lorry driver. And this possibility is obscured from view owing to the influence of the morality system not only on us as moral agents but specifically on moral philosophy. This would support the strong interpretive claim, i.e., that what Williams describes as the features of agent-regret just are the structure of a broader conception of remorse; the basis for a conception of remorse after it has been freed from the strictures of the morality system. If this is correct, it is not merely a verbal point but helps moral philosophers to acknowledge that we have this form of moral responsibility that, for Williams, is not to be integrated into our understanding of agency and first-personal responsibility as, in Enoch's words, "penumbral" to an otherwise "core centre" of agency (which is the will, the site of control). Rather, for Williams, the fact of this form of moral responsibility reveals a deeper truth concerning morality, the first-personal, agency, and responsibility: "that if one attaches importance to the sense of what one is in terms of what one has done and what in the world one is responsible for, one must accept much that makes its claim on that sense solely in virtue of its being actual" (1981b, 30).

Remorse as the purely moral, first-personal characteristic reaction in the morality system carries with it many burdensome connotations and implications for anyone interested in moving out of the system. My suggestion is that, at the very least, agent-regret should be interpreted as part of a move away from the narrow conception of remorse found in the morality system. I want to suggest that "agent-regret" is a term coined as a tool or device to get away from the baggage of the morality system and towards a different, broader, understanding of a very important ethical phenomenon: our responsibility for what we cause. If this is the case, to pursue what Williams has to say about agent-regret, I think, will help us to imagine, and construct, what form remorse might take beyond the morality system.

CONCLUSION

In the introduction to this thesis, I claimed that what makes remorse a *moral* emotion is the distinctive relation it bears to those harmed by one's actions and their consequences. As we saw, this claim finds expression in Rawls' "association guilt", and in Taylor's and Gaita's conceptions of remorse. This claim also helped to reveal a key insight: that remorse is relational. Namely, remorse exhibits at once an outward and inward trajectory because it relates the subject and object of a harmful or wrongful action. Moreover, this claim served to anchor the central objection against MS-remorse, as well as Taylor's conception of guilt.

As I argued across the chapters, if we accept and *prioritise* the idea that remorse has a special connection with those harmed, then this raises doubts about the plausibility of the claim that remorse is restricted to the voluntary. In one sense, this is what I take Williams to have been attempting to show by making "agent-regret" such a capacious category—the very fact that one has harmed another, as in the lorry driver case, may be sufficient for intelligible, coherent and reasonable remorse. This is an expression of what Williams says is "the utterly familiar fact that what has happened to others through our agency can have its own authority over our feelings, though we brought it about involuntarily" (1993, 92). Given the incitement to philosophical inquiry I take this to present, I want to claim that our further explorations for a conception or form of remorse, freed from the distortions of the morality system, ought to begin with the idea that *remorse is a first-personal expression of responsibility for harm done to others*.

Importantly, this formulation of remorse is neutral with respect to the voluntariness or intentionality of the harmful action and its consequences. That is, we can consider how or whether or to what extent the voluntary can be built-in or adjusted or explained in relation to this formulation (which, depending on our answers, may dissolve the formulation). Presently, there are at least two currently identifiable routes to take with respect to this issue. In one direction, we could follow many philosophers who now accept a broader notion of the voluntary as extending to culpable ignorance and take this notion to restrict

remorse or guilt. Alternatively, we could accept the much broader scope of remorse or guilt as extending to non-culpable ignorance that both Williams and Gaita advocate. This broader scope is focused on the distinction between what I have done and what I have not done, and not strictly on the distinction between the voluntary and involuntary. Neither philosopher is claiming that the voluntary or the intentional is unimportant in this connection. However, both Williams and Gaita claim that when one feels remorse or guilt for a harmful action that was voluntary or intentional, in terms of moral psychology, the appropriate place to look for the importance of the voluntary is *shame*.

Through the notion of a prospective genealogy, I have questioned an existing stronghold conception of remorse, and developed grounds for a broader conception of remorse than is currently standardly available in moral philosophy. In an important sense, I have raised more questions than I have provided answers. In Chapters 1 and 4, for example, we encountered the question of whether remorse is best understood as the first-personal mirror of blame or as bearing an asymmetrical relation with blame. Relatedly, in Chapter 2 we encountered the question of whether remorse has a necessary connection with culpability. In Chapter 3, we encountered the questions of what the relations between harms and wrongdoing amount to, and whether intelligible remorse conditions what we can see as wrongdoing. I take it that raising questions, as I have done in this thesis, follows from an attempt to foreground an emotion that has so far been philosophically neglected. For, surely, part of what it is to argue for the need for philosophical inquiry is first to uncover what questions and puzzles there are to ask or to consider.

I have shown that moral philosophers with an interest in moral psychology ought to consider more closely what is at stake in the rich and underexplored territory of first-personal expressions of responsibility. And, although it is too early to say how the prospective genealogy of remorse would end, on my reading of agent-regret, and the various constructive materials offered throughout this thesis, we can see how it might end. Namely, we can see how remorse has been pressurized into a highly confined form, and how, once those pressures are lifted, we can discover a broader, arguably less moralistic, conception of remorse

REFERENCES

- Baron, Marcia, "Remorse and Agent-Regret", *Midwest Studies in Philosophy*, 13 (1988), 259-281.
- Chappell, Timothy, *Knowing What To Do: Imagination, Virtue, and Platonism in Ethics*, (Oxford: Oxford University Press, 2014).
- Cordner, Christopher, "Guilt, Remorse and Victims", *Philosophical Investigations*, 30 (2007), 337-362.
- _____, "Remorse and Moral Identity", in *Practical Identity and Narrative Agency*, Kim Atkino and Catriona Mackenzie (eds.), (Abingdon: Routledge, 2008), 232-251.
- Davidson, Donald, "Hume's cognitive theory of pride", *Journal of Philosophy*, 73 (1976), 744-757.
- de Sousa, Ronald, *The Rationality of Emotion* (Cambridge: MIT Press, 1987).
- de Wijze, Stephen, "Tragic-Remorse–The Anguish of Dirty Hands", *Ethical Theory and Moral Practice*, 7 (2004), 453-471.
- Darwall, Stephen, *The Second Person Standpoint: Morality, Respect and Accountability*, (Cambridge: Harvard University Press, 2006).
- Deigh, John, "Review: Pride, Shame and Guilt: Emotions of Self-Assessment by Gabriele Taylor", *Ethics*, 98 (1988), 391-393.
- _____, "Cognitivism in the Theory of Emotions", *Ethics*, 104 (1994), 824-854.
- Doris, J. M., and The Moral Psychology Research Group (eds.), *The Moral Psychology Handbook*, (Oxford: Oxford University Press, 2010).
- Doris, J. M. and Stich, S. P., "As a Matter of Fact: Empirical Perspectives on Ethics", in *The Oxford Handbook of Contemporary Philosophy*, F. Jackson and M. Smith (eds.), (Oxford: Oxford University Press, 2005), 114-152.
- Duff, Antony, "Good and Evil and the criminal law", in *Philosophy, Ethics and A Common Humanity: Essays in Honour of Raimond Gaita*, Christopher Cordner (ed.) (Abingdon: Routledge, 2011), 68-81.
- Enoch, David, "Being Responsible, Taking Responsibility, and Penumbral Agency", in *Luck, Value and Commitment: Themes from the Ethics of Bernard Williams*, Ulrike Heuer and Gerald Lang (eds.), (Oxford: Oxford University Press, 2012), 95- 132.
- Foot, Philippa, "Moral Realism and Moral Dilemma", *The Journal of Philosophy*, 80 (1983), 379–398.
- _____, *Natural Goodness*, (Oxford: Oxford University Press, 2001).
- Fricker, Miranda, "Boundaries of Moral Philosophy", in *Reading Ethics: Selected Texts with Interactive Commentary*, Miranda Fricker and Samuel Guttenplan (eds.), (Chichester: Blackwell Publishing, 2009), 266-316.
- _____, "What's the Point of Blame? A Paradigm Based Explanation", *Noûs*, (May 2014), 1-19.

- Gaita, Raimond, *A Common Humanity: Thinking About Love and Truth and Justice*, 2nd edition, (Abingdon: Routledge, 2000).
- _____*Good and Evil: An Absolute Conception*, 2nd edition, (Abingdon: Routledge, 2004).
- _____*After Romulus*, (Melbourne: Text Publishing, 2011).
- Goldie, Peter, *The Emotions: a Philosophical Exploration*, (Oxford: Oxford University Press, 2000).
- Greenspan, Patricia, "Moral Dilemmas and Guilt", *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 43 (1983), 117-125.
- _____*Emotions and Reasons: an Inquiry into Emotional Justification*, (New York: Routledge, Chapman and Hall, 1988).
- _____"Subjective Guilt and Responsibility", *Mind*, 101 (1992), 287-303.
- _____*Practical Guilt: Moral Dilemmas, Emotions, and Social Norms*, (Oxford: Oxford University Press, 1995).
- Hursthouse, Rosalind, *On Virtue Ethics*, (Oxford: Oxford University Press, 1999).
- Hyman, John, "Voluntariness and Choice", *The Philosophical Quarterly*, 63 (2013), 683-708.
- Jacobson, Daniel, "Regret, Agency, and Error", in *Oxford Studies in Agency and Responsibility: Volume 1*, David Shoemaker (ed.), (Oxford: Oxford University Press, 2013), 95-125.
- Jenkins, Mark P., *Bernard Williams (Philosophy Now Series)*, (Montreal: McGill-Queen's University Press, 2006).
- Kenny, Anthony, *Action, Emotion and Will*, (London: Routledge, 1963).
- Leiter, Brian, "Nietzsche and the Morality Critics", *Ethics*, 107 (1997), 250-285.
- Louden, Robert B., "The Critique of the Morality System", in *Bernard Williams: Contemporary Philosophy in Focus*, Alan Thomas (ed.), (Cambridge: Cambridge University Press, 2007), 104-134.
- McGinn, Marie, *The Routledge Guidebook to Wittgenstein's Philosophical Investigations*, (Abingdon: Routledge, 2013).
- MacIntyre, Alasdair, "Moral Dilemmas", *Philosophical and Phenomenological Research*, 50 (1990), 367-382.
- Mason, H.E., (ed.), *Moral Dilemmas and Moral Theory*, (New York: Oxford University Press, 1996).
- Marcus, Ruth Barcon, "Moral Dilemmas and Consistency", *The Journal of Philosophy*, 77 (1980), 121-136.
- Moran, Richard, *Authority and Estrangement: An Essay on Self-Knowledge*, (Woodstock: Princeton University Press, 2001).
- Nagel, Thomas, "Moral Luck", in *Mortal Questions*, (Cambridge: Cambridge University Press, 1979), 24-38.

- Nietzsche, Friedrich, *Beyond Good and Evil*, Walter Kaufmann (trans.), (New York: Random House, 1966).
- Nussbaum, Martha, *Love's Knowledge*, (Oxford: Oxford University Press, 1990).
- Phillips, D.Z., and Price, H.S., "Remorse Without Repudiation", *Analysis*, 28 (1967), 18-20.
- Proeve, M., and Tudor, S., *Remorse: Psychological and Jurisprudential Perspectives*, (Surrey: Ashgate Publishing Ltd., 2010).
- Rawls, John, *A Theory of Justice: Revised Edition*, (Cambridge: Belknap Press, 1999).
- Raz, Joseph, "Agency and Luck", in *Luck, Value and Commitment: Themes from the Ethics of Bernard Williams*, Ulrike Heuer and Gerald Lang (eds.), (Oxford: Oxford University Press, 2012), 133-161.
- Rorty, Amélie, "Agent Regret", in *Explaining Emotions*, Amélie Rorty (ed.), (London: University of California Press, 1980), 489-506.
- Russell, Paul, "Responsibility, naturalism, and 'The Morality System'", in *Oxford Studies in Agency and Responsibility: Volume 1*, David Shoemaker (ed.), (Oxford: Oxford University Press, 2013), 184-204.
- Scheler, Max, "Repentance and Rebirth", in *On the Eternal in Man*, Bernard Noble (trans.), (Hamden: Archon Books, 1972), 35-65.
- Sinnott-Armstrong, Walter, *Moral Dilemmas*, (Oxford: Basil Blackwell, 1988).
- Strawson, P.F., "Freedom and Resentment", in *Freedom and Resentment and Other Essays*, (Abingdon: Routledge, 2008), 1-28.
- Taylor, Charles, "A Most Peculiar Institution", in *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams*, J.E. Altham and Ross Harrison (eds.), (Cambridge: Cambridge University Press, 1995), 132-155.
- Taylor, Gabriele, *Pride, Shame and Guilt: Emotions of Self-Assessment*, (Oxford: Clarendon Press, 1985).
- Thomson, Judith Jarvis, "Morality and Bad Luck", in *Moral Luck*, Daniel Statman (ed.), (New York: State University of New York Press, 1993), 195-216.
- Tudor, Steven, *Compassion and Remorse: Acknowledging the Suffering Other*, (Leuven: Peeters Publishers, 2001).
- Velleman, J. David, "Don't Worry, Feel Guilty", in *Philosophy and the Emotions: Royal Institute of Philosophy Supplement* (vol. 52), Anthony Hatzimoysis (ed.), (Cambridge: Cambridge University Press, 2003), 235-248.
- Wallace, R. Jay, *Responsibility and the Moral Sentiments* (Cambridge: Harvard University Press, 1994).
- _____ *The View From Here* (Oxford: Oxford University Press, 2013).
- Williams, Bernard, "Ethical consistency", in *Problems of the Self: Philosophical Papers 1956-1972* (Cambridge: Cambridge University Press, 1973a), 166-186.

- _____ “Morality and the emotions”, in *Problems of the Self: Philosophical Papers 1956-1972* (Cambridge: Cambridge University Press, 1973b), 207-229.
- _____ “Persons, character and morality”, in *Moral Luck: Philosophical Papers 1973-1980* (Cambridge: Cambridge University Press, 1981a), 1-19.
- _____ “Moral luck”, in *Moral Luck: Philosophical Papers 1973-1980* (Cambridge: Cambridge University Press, 1981b), 20-39.
- _____ “Politics and moral character”, in *Moral Luck: Philosophical Papers 1973-1980* (Cambridge: Cambridge University Press, 1981c), 54-70.
- _____ “Conflicts of values”, in *Moral Luck: Philosophical Papers 1973-1980* (Cambridge: Cambridge University Press, 1981d), 71-82.
- _____ *Shame and Necessity* (Berkeley: University of California Press, 1993a).
- _____ *Morality: An Introduction to Ethics* (Cambridge: Cambridge University Press, 1993b).
- _____ “How free does the will need to be?”, in *Making Sense of Humanity and Other Philosophical Papers* (Cambridge: Cambridge University Press, 1995a), 3-21.
- _____ “Voluntary acts and responsible agents”, in *Making Sense of Humanity and Other Philosophical Papers* (Cambridge: Cambridge University Press, 1995b), 22-34.
- _____ “Internal reasons and the obscurity of blame”, in *Making Sense of Humanity and Other Philosophical Papers* (Cambridge: Cambridge University Press, 1995c), 35-45.
- _____ “Nietzsche’s minimalist moral psychology”, in *Making Sense of Humanity and Other Philosophical Papers* (Cambridge: Cambridge University Press, 1995d), 65-76.
- _____ “Moral luck: a postscript”, in *Making Sense of Humanity and Other Philosophical Papers* (Cambridge: Cambridge University Press, 1995e), 241-247.
- _____ *Truth and Truthfulness: An Essay in Genealogy*, (Princeton: Princeton University Press, 2002).
- _____ *Ethics and the Limits of Philosophy* (Abingdon: Routledge, 2006a).
- _____ “Moral Responsibility and Political Freedom”, in *Philosophy as a Humanistic Discipline* (Princeton: Princeton University Press, 2006b), 119-125.
- _____ “Philosophy as a Humanistic Discipline”, in *Philosophy as a Humanistic Discipline* (Princeton: Princeton University Press, 2006c), 180-199.
- _____ “Understanding Homer: Literature, History and Ideal Anthropology”, in *The Sense of the Past* (Princeton: Princeton University Press, 2006d), 60-68.
- _____ “Review of ‘The English Moralists’”, in *Essays and Reviews* (Oxford: Princeton University Press, 2014), 53-54.

Wittgenstein, Ludwig, *Philosophical Investigations*, 3rd edition, G.E.M. Anscombe (trans.), (Oxford: Blackwell Publishers, 2001).