

HEIDY MERISTE

The Emotion of Guilt:  
Integrating Cases Without  
Perceived Wrongdoing





**HEIDY MERISTE**

The Emotion of Guilt: Integrating Cases  
Without Perceived Wrongdoing



UNIVERSITY OF TARTU

Press

Faculty of Arts and Humanities, Institute of Philosophy and Semiotics

This dissertation has been accepted for defence of the degree of Doctor of Philosophy (PhD) in Philosophy by the Council of the Institute of Philosophy and Semiotics, University of Tartu, on the 16<sup>th</sup> of May 2025.

Supervisor: Prof Margit Sutrop (University of Tartu)

Opponent: Prof Edward Harcourt (University of Oxford)

Defence: This dissertation will be defended on the 19<sup>th</sup> of August 2025 at 14.15 in Ülikooli 18-128, Tartu

The research was supported by the Centre of Excellence in Estonian Studies (European Union, European Regional Development Fund); Doctoral Studies and Internationalisation Programme DoRa (European Union, European Social Fund; through the Archimedes Foundation); the project “INspiring and ANchoring TrUst in Science” (IANUS), grant nr 101058158 (European Union, Horizon Union); the project “Promoting Integrity as an Integral Dimension of Excellence in Research” (PRINTEGER), grant nr 665926 (European Union, Horizon Europe); the project “Disagreements Philosophical Analysis”, IUT 20-5 (Estonian Ministry of Education and Research); and the project “Speech that Silences Speech: Mechanisms and Solutions”, PUT632 (Estonian Ministry of Education and Research).



European Union  
European Regional  
Development Fund



Investing  
in your future

ISSN 1406-9520 (print)  
ISBN 978-9916-27-944-1 (print)  
ISSN 2806-2515 (pdf)  
ISBN 978-9916-27-945-8 (pdf)

Copyright: Heidi Meriste, 2025

University of Tartu Press  
[www.tyk.ee](http://www.tyk.ee)

# CONTENTS

ACKNOWLEDGEMENTS .....	7
1. INTRODUCTION.....	8
1.1. The main question of this dissertation and its importance .....	8
1.1.1. The main question .....	8
1.1.2. The practical importance of the question .....	9
1.1.3. The theoretical importance of the question .....	12
1.2. Guilt and contemporary analytic philosophy .....	13
1.2.1. On the neglect of guilt in contemporary analytic philosophy ..	13
1.2.2. Why do we need a philosophical account of guilt at all? .....	18
1.3. The account of guilt defended in this dissertation .....	21
1.3.1. The main ideas.....	21
1.3.2. Methodology .....	25
1.3.3. Chapter Outline .....	27
2. GUILT AND MORAL WRONGDOING.....	31
2.1. The traditional account of guilt.....	31
2.1.1. The cognitivist aspect.....	31
2.1.2. The moral wrongdoing aspect .....	33
2.2. The problem cases.....	37
2.2.1. Collective guilt .....	38
2.2.2. Self-disciplinary guilt .....	39
2.2.3. Guilt for merely causal harm .....	42
2.2.4. Survivor guilt.....	46
2.3. The traditionalist response to the problem cases.....	51
2.3.1. Attribution of irrational judgments or thoughts.....	52
2.3.2. Denying the status of guilt.....	53
2.4. How to move forward? .....	58
3. GUILT FOR HARM TO OTHERS .....	65
3.1. The interpersonal account and the non-moralisation of guilt.....	66
3.1.1. The interpersonal account of guilt.....	66
3.1.2. The nonmoralisation of emotions .....	69
3.2. In defence of a moralised view of harm.....	79
3.2.1. Making room for feeling sorry without guilt.....	81
3.2.2. Harm and the bad .....	83
3.2.3. Undeserved harm.....	85
3.2.4. In defence of desert as a moral matter.....	90
4. SURVIVOR GUILT .....	103
4.1. At-the-expense guilt and harm to others .....	105
4.2. Normative vulnerability and anxiety about resentment .....	110

4.2.1. Velleman on guilt as anxiety about resentment.....	111
4.2.2. MacKenzie and Zhao on guilt as a feeling of normative vulnerability.....	123
4.3. Survivor guilt and comparative injustice .....	125
4.3.1. Comparative injustice as the core of survivor guilt.....	126
4.3.2. Comparative injustice as the foundation of more complex forms of guilt.....	129
5. THEORIES OF GUILT .....	133
5.1. The value-based account of guilt .....	133
5.1.1. The elements of the value-based account of guilt .....	134
5.1.2. Alternative phrasings: moral footprint and violating the moral order .....	146
5.1.3. Further applications.....	148
5.2. Comparison with other accounts of guilt .....	151
5.2.1. A midway between the traditional and interpersonal accounts of guilt .....	151
5.2.2. Comparison with other value-based accounts of guilt.....	154
5.3. Normative ethics and the limitations of the value-based account.....	158
5.3.1. On the moral theories that the value-based account does cover .	159
5.3.2. On the moral theories that the value-based account does not cover.....	164
5.3.3. On the prospects of a unified account of guilt.....	169
6. CONCLUSION .....	172
BIBLIOGRAPHY .....	177
SUMMARY IN ESTONIAN .....	188
CURRICULUM VITAE .....	192
ELULOOKIRJELDUS.....	195

## ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude to everyone who has been there with me throughout the writing of this dissertation, of whom I can mention only a fraction here.

Most importantly, I would like to thank my supervisor, Margit Sutrop, who has always been there for me throughout my studies. Her advice and encouragement, her kindness and patience have helped me come a long way. My research has also greatly benefited from the expertise of Fabrice Teroni, under whose supervision I spent two months in Geneva back in 2018, and from Francesco Orsi's helpful remarks on normative theory. I am also grateful to Jaanus Sooväli for helping me keep my horizons broad with respect to the Nietzschean-Freudian approach to guilt, and to Bruno Mölder for helping me become more attentive to broader matters of theoretical philosophy.

I would like to thank the reviewers of this dissertation, Edward Harcourt and Timo Airaksinen, whose comments helped to clarify and improve this work.

I am indebted to Corey Maley, whose invitation to contribute a chapter to "The Moral Psychology of Guilt" (Cokelet & Maley, 2019) collection played a pivotal role in shaping my path towards focusing my research on guilt feelings in particular. I do not think that this particular thesis would have been written if our paths had not crossed.

I am grateful to the Department of Philosophy and the Chair of Practical Philosophy for providing me with a stimulating academic environment and for the chance to teach courses on both emotions and ethics. My interest in guilt was greatly facilitated by the opportunity to teach the course "Guilt: Moral and Conceptual Aspects" back in 2018, and my regular course "The main topics of ethics" has ensured that I have continually kept myself in shape on the front of normative ethics. It is said that teaching is the best way to learn, and I could not agree more.

My philosophical work on morality and guilt has also been greatly facilitated by my employment in the Centre for Ethics, where I have had the opportunity to work on questions of ethics throughout many projects and endeavours. I have been lucky to be surrounded by the most wonderful and supportive colleagues. Given my duties in the Library of the Department of Philosophy and the Centre for Ethics, my special thanks goes out to my fellow librarians Laura Lilles-Heinsar and Eva Lepik (the latter from our sister library of the Department of Semiotics), who have kindly put up with the fact that academic life involves a changing schedule of responsibilities, and who have had to put in extra hours all the times I have been on study leave.

Last but not least, I would also like to thank the coordinator of the Institute of Philosophy and Semiotics, Ruth Jürjo, who has always been there for me with regard to understanding the peculiarities of the procedural matters of doctoral studies, and who has never left me with unanswered questions or without a note of encouragement (or a little candy).

# 1. INTRODUCTION

## 1.1. The main question of this dissertation and its importance

### 1.1.1. The main question

The emotional phenomenon of guilt is traditionally defined as a painful feeling about having done something morally wrong.<sup>1</sup> Since the emphasis is on the agent having misbehaved, guilt feelings are considered appropriate in case of all kinds of moral misdemeanours like lying, cheating, stealing, and more serious crimes like murder. Yet it is also common to report feelings of guilt in the absence of perceived wrongdoing. There are two most noteworthy types of cases. First are the cases where the subject is just causally rather than morally responsible for harming other people. For example, drivers who, through no fault of their own, have killed another person in a car accident (Gregory, 2017; Zhao, 2020), or people who cannot reciprocate another person's romantic interest and thereby become reluctant agents of their heartbreak (Baumeister & Wotman, 1994). Secondly, there are cases of unfair benefits, most well known in the form of survivor guilt. In those cases, the subject need not even see oneself as causally responsible for other people's fate, yet still feels bad for being more fortunate. Both types of cases have been widely discussed in both psychological research and also less formal avenues such as media interviews, blog posts and autobiographies.

The proponents of the traditional approach would have to argue that those are not cases of true guilt or must nevertheless involve some irrational judgment of moral culpability. Siding with many others, I find it uncharitable to doubt people's self-ascriptions of emotional states or to suspect that they must be in the grips of some "profound linguistic and conceptual confusion" (Fischer 2017, 188, see also Morris 1987, 221). Especially in the case of an emotion like guilt that does not seem to have a particularly pronounced physiological profile (like the heart beating faster in case of fear, for example), I find it hard to see how we should proceed with our research at all if we do not give weight to self-report. So, inclined to take people's self-reports seriously, I am interested in explicating what kind of emotional experience people capture when they refer to cases without perceived wrongdoing and exploring whether our conception of guilt could be broadened to include such a wider range of cases. Rather than just brushing those cases aside as 'not guilt' or explaining them as some bizarre anomalies, I am interested in the prospects of integrating them as full-fledged members of the family of guilt that can be a meaningful, albeit tragic, part of people's lives.

---

<sup>1</sup> Among others, the proponents of this view include but are not limited to the following authors: John Rawls (2005), R. E. Lamb (1983), Walter Sinnott-Armstrong (2005), John Deigh (1982, 1999, 2019), Patricia Greenspan (1992, 1995), Martha Nussbaum (2016), Andrew Ingram (2017, 2020).

Since I take it that some will be reluctant to extend the label of ‘guilt’ to those cases, I would like to emphasise from the outset that even though I side with the more liberal use of the word ‘guilt’ and frame my project as a quest for a wider theory of guilt, the most fundamental question for me is not whether we should necessarily label those cases ‘guilt’ but whether those cases form a unique genus with the more traditional instances of guilt: is the emotional experience in cases without perceived wrongdoing structurally similar to that of more traditional instances of guilt? Or is it just the case that people engage in folk psychological loose talk when they extend the label of guilt to cases without perceived wrongdoing—something that should not be taken overly seriously and that should not find a place in a more rigorous discourse on guilt?

Given all the social psychological research on guilt without perceived wrongdoing, one might expect that philosophers, too, have paid more attention to the possibility of integrating the cases without perceived wrongdoing into a more general theory of guilt. Unfortunately, this has not been the case. With the exception of Michael Zhao and John Deigh (the latter of whom is more focused on why we should *not* integrate those cases under a theory of guilt), the question has not really received much attention. The standard philosophical approach is the one that defines guilt in terms of perceived wrongdoing, and according to that approach, guilt without perceived wrongdoing is just a conceptual impossibility—to the extent that the abovementioned cases are truly instances of guilt they must involve some covert irrational judgment about wrongdoing, or else we must be dealing with something other than guilt. Of course, there is no logical necessity that would stop the proponents of the standard view of guilt from developing a separate theory of some different label to account for cases without wrongdoing—an endeavour I would wholeheartedly salute (I will have more to say about different labelling practices later on). Unfortunately, in practice, adopting the standard approach usually results in cases without perceived wrongdoing being neglected from a more rigorous philosophical study. Instead of receiving positive treatment, those cases are just brushed aside as ‘not guilt’.

### 1.1.2. The practical importance of the question

I will start with the practical importance of the question because, beyond all the implications that it has for field-specific theoretical debates, the ultimate reason why we need an analysis of guilt is a practical matter—it can help us better understand both ourselves and other people. In fact, even though cases without perceived wrongdoing have not received much philosophical attention, it could be argued that those are actually the cases where a more detailed inquiry would be the most helpful. People who experience guilt without wrongdoing often find themselves in rather tragic circumstances, and the fact that their emotional experience does not accord well with common presumptions about guilt makes things no easier. It can make it harder for them to make sense of their emotion, and it also makes it harder for their family and friends, therapists and priests to be more understanding and empathetic. While an account of guilt without wrongdoing

may seem utterly helpless in undoing the tragedies of the past, it can nevertheless provide important support for coping with the past and moving on.

Let me first focus on how an account of guilt can be useful for the subject who experiences such guilt. As already mentioned, it can be confusing to feel guilt without wrongdoing because it does not accord well with the dominant discourse on guilt that ties guilt with a sense of moral culpability. The importance of finding “a way to connect the dots between” what happened and the guilt feelings that followed (Peters, 2023, p. 38) has been well captured by David W. Peters, who caused the death of a motorcyclist in an automobile accident.

I did not have the language for who I was or what I had done. My crash drifted into the murky moral area I knew nothing about. (Peters, 2023, p. 36)

I wasn't able to go to confession for my accident. Since it was technically an “accident,” why should I feel guilty? Society and even my own religious faith assigned no guilt to me, so confessing it as a “sin” didn't make much sense. (Peters, 2023, p. 38)

Peters stresses both that being able to conceptually make sense of one's experience would be a relief on its own, but also that it would allow one to more freely talk to others about how they feel and to seek help.

As also noted in Alice Gregory's *The New Yorker* piece “The Sorrow and the Shame of the Accidental Killer” (2017), in cases of guilt for merely causal harm like motor vehicle accidents, the lack of resources about coping with the aftermath is especially great:

There are self-help books written for seemingly every aberration of human experience: for alcoholics and opiate abusers; for widows, rape victims, gambling addicts, and anorexics; for the parents of children with disabilities; for sufferers of acne and shopping compulsions; for cancer survivors, asexuals, and people who just aren't that happy and don't know why. But there are no self-help books for anyone who has accidentally killed another person. An exhaustive search yielded no research on such people, and nothing in the way of therapeutic protocols, publicly listed support groups, or therapists who specialize in their treatment. (Gregory, 2017)

Some time has passed since this article, and there have been improvements. Much of those improvements have been led by the Hyacinth Fellowship (the former Accidental Impacts), a California-based non-profit corporation, which is the only organisation worldwide that is dedicated to helping all those who have accidentally killed another person. In fact, David W. Peters, whom I quoted just earlier, was associated with the organisation as an author and, inspired by this very fragment by Gregory, wrote the first self-help book “Accidental: Rebuilding a Life after Taking One” (2023). This was later followed by “Accidental Killing: A Survivor's Handbook” (2024) by the founder, Maryann Jacobi Gray, and the current board president, Chris Yaw. Gray has also published a review article “Unintentional Killing: A Neglected Trauma” (2021).

There is no doubt that these are important improvements that provide much-needed help for a great number of people but the topic nevertheless remains relatively little researched, especially through a philosophical and also a more secular angle—two of the above-mentioned authors, Peters and Yaw, are priests and tend to focus on the more religious aspect of the experience. While this dissertation is not a self-help book but a philosophical treatise written in academic language, it is nevertheless my hope that the ideas presented in this dissertation could be of help to those who need them the most—that the account I offer could provide for at least some people a useful starting point to put their experience into words, that it would help them to disentangle the mess inside and connect the dots between what happened and the feeling that followed. As such, it will also help to promote higher emotional granularity. Emotional granularity is a term coined by psychologist Lisa Feldman Barrett, which refers to the ability to make more fine-grained distinctions within our emotional life. Just like painters who can make fine-grained distinctions within colours are better equipped to work with colour, people who make more fine-grained distinctions within their emotions are shown to be “better calibrated to cope with varied circumstances” and more flexible when regulating their emotions (Barrett, 2017, p. 182).

So far, I have focused on how an account of guilt without wrongdoing can support people who experience guilt without wrongdoing to make better sense of this inherently confusing emotional experience. Now, I want to highlight why an account of guilt without wrongdoing is also important for all others, who may come into contact with people who feel guilt without perceived wrongdoing—especially their family members and friends, therapists and priests. It is important because it can help us all better understand those who go through guilt without wrongdoing and thereby also support them in their process of healing.

For people who feel guilt, it can be difficult enough to overcome the silence and share their experience, but even if they do, they often end up being misunderstood. It is common to try to comfort them with phrases like “It was just an accident” or “It was not your fault”. Such phrases are usually well-meant and do offer assurance in the sense that it is good to know that at least one is not being blamed for the situation. But it seems that sometimes, at least, such words of comfort are perceived as missing the point. People have even been reported to grow impatient at hearing such attempts to comfort them, crying out: “Yes, it was an accident, [---] but, at the end of the day, I *hit* him, I took *his* life”, “I hit him! Why does nobody understand this?” (Gregory, 2017; see also Zhao, 2020, p. 287). It can be frustrating if one’s sincere report of guilt feelings is repeatedly met with a “But it’s not your fault”. Peters also reports that when, many years later, he finally mustered the courage to talk about the accident and his guilt feelings to a priest, he was simply told that he did not need to confess because it was just an accident (Peters, 2023, p. 40). Again, there is no doubt that those words were well meant, but not only did the priest fail to understand Peters’s emotions, he also used those words as a conversation stopper because he was unable to see how there could be anything further to say about the matter—even though Peters really

needed to talk to someone and was hoping to get some spiritual guidance about how to cope with such an existentially shattering experience.

Perhaps this happens less often, but especially if one's confession of guilt feelings is repeatedly met with "But it's not your fault", such words can also be potentially perceived as having a discrediting effect because they imply that one's emotional reaction is somehow inadequate, possibly because they irrationally believe that what happened was nevertheless their fault. Or else, because the person is perhaps misidentifying their emotion—mistaking something else for guilt. Even if the traditional approach that ties guilt feelings with moral wrongdoing may look good on paper, having the corresponding responses reflected in a real-life interaction does not make them look particularly attractive. Of course, we might have reasons to refrain from showing any doubt in traumatised people's emotional reactions, even if those reactions are indeed somehow irrational or confused. But if there was indeed a good rationale to consider the abovementioned feelings as a form of guilt, then the situation where the conversation partner just thinks to themselves that the person who reports guilt feelings must be confused and does not openly object to her would hardly contribute to any deeper kind of mutual understanding either.

### 1.1.3. The theoretical importance of the question

Next, let us move on to the importance that the question of guilt without moral wrongdoing has on the theoretical landscape of guilt. As philosophers have not really paid much attention to cases of guilt without perceived wrongdoing, their theoretical importance becomes more evident in debates in social psychology. Those cases reflect the fundamental clash between intrapsychic and interpersonal theories of guilt—a clash that has been most distinctly carved out in the seminal article "Guilt: An Interpersonal Approach" by Roy Baumeister and his colleagues (1994). The traditional account of guilt in terms of perceived wrongdoing is a version of what they label as the intrapsychic account of guilt—it involves the subject evaluating themselves as having done something morally wrong.<sup>2</sup> As already explained, such an account is not particularly hospitable to cases without perceived wrongdoing. For this reason, Baumeister and his colleagues argue that we should instead adopt an interpersonal account of guilt, according to which guilt is not essentially about moral wrongdoing but more fundamentally about harming other people.

---

<sup>2</sup> There are also other accounts that could be labelled intrapsychic—for example, Sigmund Freud's (1962) account of guilt as the ego's fear of a superego's harsh response. This dissertation does not deal with fear-based accounts of guilt because, even though fear may play some role in the development of guilt feelings, mature guilt feelings are widely conceived as reflecting non-egoistic motivation—concern about having done the wrong thing, not being good to others. Fear, however, is linked to egoistic motivation and the avoidance of punishment. As such, I am concerned with the traditional account of guilt in terms of moral wrongdoing as one of the strongest versions of the intrapsychic approach and focused only on that.

The cases of merely causal harm and guilt for unfair benefits are the central cases that the intrapsychic and interpersonal theories of guilt disagree upon. So, how we make sense of those cases is of central importance for the theory formation on guilt. As already said, the traditional approach as the central example of an intrapsychic approach, suggests that those cases should not count as guilt at all, whereas the interpersonal approach implicitly suggests that those are the cases where the true essence of guilt—the focus on harm—becomes the most visible because those cases are purified from the somewhat distracting element of moral culpability.<sup>3</sup> Yet if we say that guilt is solely about causing harm to other people, then this may have downsides of its own. For example, it is heavily influenced by the harm-based ethics of the WEIRD (Western, Educated, Industrialised, Rich and Democratic) cultures, and does not seem to do justice to the experience of religious guilt, which is often viewed as a matter of not playing God (Mancini & Gangemi, 2021, p. 5).

Should it be the case that both approaches to guilt, the intrapsychic and the interpersonal one, are ill-fitted to account for all instances of guilt, the question arises: are there prospects of unification at all? Is there a third way to unify the sphere of guilt, or should we just say that the sphere of guilt fragments into different types of guilt that cannot be brought under a common denominator? These are all questions that this dissertation will address.

## **1.2. Guilt and contemporary analytic philosophy**

### **1.2.1. On the neglect of guilt in contemporary analytic philosophy**

One of the aims of this dissertation is to enrich those interdisciplinary debates from a philosophical angle. As already said, philosophy seems somewhat out of sync with those developments in social psychology. In what follows, I will first offer some hypotheses about how it has come about that guilt is a relatively neglected emotion in philosophical theorisation and then go on to suggest why philosophical theorisation might be especially well suited to help us resolve some of the mysteries of guilt without wrongdoing.

One of the reasons why, during recent decades at least, guilt has been a relatively neglected emotion is the fact that it is shame that has stolen the spotlight. In part, this reflects a more general shift that has taken place in many sciences like medical and psychiatric sciences, literary criticism and philosophy, and can be described as a movement “away from the “moral” concept of guilt in favour of the ethically different or “freer” concept of shame” (Leys, 2007, p. 7). A prominent philosophical monograph that reflects this shift is Bernard Williams’s “Shame and Necessity” (1993). The reasons behind such broad shifts are, of

---

<sup>3</sup> This is not to say that the interpersonal approach could not consider more traditional cases that do involve moral wrongdoing as prototypical cases. According to the interpersonal approach, those cases are cases of harm as well, but, at least in the frames of the intrapsychic approach, the harm element gets overshadowed by the element of moral culpability.

course, complex, but perhaps it is understandable that in the context of guilt cultures (I am focused on the philosophical debates within the Anglo-American tradition of analytic philosophy), shame seems like a more intriguing emotion. We are used to taking the importance of guilt for granted and are perhaps also a little bit tired of a highly moralised society's burden of guilt.<sup>4</sup>

Within the more recent philosophical literature, philosophical interest in shame has, in great part, also been a counteraction to certain developments in social psychology. Much in line with what might be expected in a guilt culture, social psychologist June Tangney and her colleagues have argued that guilt is a constructive prosocial emotion, whereas shame is its destructive “ugly sister”: rather than motivating reparation, it just makes the person withdraw and feel bad (Tangney et al., 2007; Tangney & Dearing, 2003). Many philosophers have been tempted to argue against this view, so there has been much controversy around whether shame is really as ugly as it has been claimed. In addition to individual articles like Kristján Kristjánsson's “Is Shame an Ugly Emotion?” (2014), this controversy has also resulted in several important monographs published on shame—“In Defense of Shame: The Faces of an Emotion” (2012) by Julien Deonna, Fabrice Teroni and Raffaele Rodogno, and Krista Thomason's “Naked: The Dark Side of Shame and Moral Life” (2018). In contrast, the last important philosophical monograph on guilt was published in 1995—Patricia Greenspan's “Practical Guilt: Moral Dilemmas, Emotions, and Social Norms”. This is almost thirty years ago.

Even though the polemics around shame as the ugly sibling of guilt, have made guilt seem like the older and wiser, though perhaps the less interesting one of the two, it cannot be denied that to some extent, guilt has, of course, been following along in discussions of shame—if not as the main character, then at least as an important side character that provides a useful point of comparison. Since the emotions often occur in conjunction, it is important to differentiate between the two, and there are several articles that are dedicated to how exactly we should (and should not) draw the line (Teroni & Bruun, 2011; Teroni & Deonna, 2008; Miceli & Castelfranchi, 2018; Maibom, 2019).<sup>5</sup> Here are some prominent points of comparison. Guilt is exclusively related to morality, whereas the sphere of shame is wider—people can, of course, be ashamed of their immoral

---

<sup>4</sup> The idea that civilised society is burdened by guilt comes from Sigmund Freud (1962) and Friedrich Nietzsche (2007). The basic idea lies in the suggestion that since the pressures of civilised society force people to suppress their aggressive urges, those urges do not disappear but are turned inwards, against the subject themselves. The subject becomes exceedingly self-critical, and this is also reflected in one's guilty conscience, which, too, is seen as a kind of self-flagellation. See also how history professor Wilfred M. McClay (2017) elaborates on the “strange persistence of guilt” in our contemporary society.

<sup>5</sup> There are also some interesting but admittedly preliminary “speculations” (B. Williams, 1993, p. 90) in the Endnote I of Bernard Williams's book “Shame and Necessity” (B. Williams, 1993, pp. 219–223). This book, too, is a reaction to guilt culture's scepticism about shame, though his focus is more historical than conceptual: he is concerned with how the modern guilt cultures look down at the ancient Greeks' shame culture as morally immature.

character, but they are often also ashamed of their nonmoral features like physical appearance (Teroni & Bruun, 2011, pp. 233–234). Shame is felt about the self, whereas guilt is felt about a particular deed (H. B. Lewis, 1971, p. 30; Teroni & Deonna, 2008, p. 730). So, even in cases where both concern moral matters, one would feel guilt about a particular immoral deed, while they would be ashamed only if they thought that this immoral deed reflected an underlying immoral character trait. Guilt is concerned with norms, whereas shame is concerned with values—we feel guilt about the transgression of norms, whereas, in the case of shame, we fail to live up to an ideal, a particular self-relevant value (Teroni & Deonna, 2008, p. 732).

Even though there is no doubt that all these comparisons are a valuable contribution to the philosophy of emotion, there is also a sense in which focusing on guilt and shame as a pair may have a downside. First, the desire to make a neat comparison can make us tacitly assume that guilt and shame belong to the same (sub)class of emotional phenomena—that they are easily comparable, like lemons and oranges. This assumption may be wrong. Following the original article by Andrew Ortony (1987), several researchers (Elison, 2005; Maley, 2015; Maley et al., 2019) have argued that guilt does not really belong to the same family with other emotion types like shame, fear, or anger but is more akin to states like feeling abandoned (Ortony, 1987), feeling alone (Elison, 2005), or feeling betrayed (Maley et al., 2019). One of the main rationales for this idea is that guilt does not pass the feel-be test, according to which, in the case of true emotions, to feel  $x$  is the same as to be  $x$  (Ortony, 1987). There is no difference between being sad and feeling sad, being ashamed and feeling ashamed. But there is such a difference when it comes to feeling guilty, abandoned, etc. As suggested by Ortony, to count all those latter phenomena as emotions in their own right has “troublesome consequences” for the theory of emotion—it allows for an indefinite number of emotion types, which offers no prospect of coherent treatment of the domain of emotions (1987, pp. 289–290). If it is indeed correct that guilt is not an emotion in the ordinary sense of the term, then guilt and shame are not really the same kind of phenomena. As suggested by Jeff Elison’s brilliant article title “Shame and Guilt: A Hundred Years of Apples and Oranges” (2005), we might be comparing phenomena the structure of which may not lend itself well to simple comparisons. It is also worth noting that Patricia Greenspan, the author of the main philosophical monograph that is available on guilt (1995), considers guilt as a rather unique emotional phenomenon as well—she argues that it is an identificatory mechanism that, through identification with the victim or someone “of the order of a judge”, can include other emotions like shame as part of it. This, too, makes guilt an oddball among all the other emotion types, shame included. In addition to the fact that the questionable status of guilt as an emotion obscures comparisons with shame, it is probably also one of the reasons why guilt is not the first choice for researchers working on the philosophy of emotion.

All the above went to show that comparisons between guilt and shame can make us tacitly assume that they are more similar to each other than we might be warranted in thinking. The second point I want to emphasise is that even if guilt

and shame would indeed belong to the same (sub)class of emotional phenomena, the desire to draw a sharp contrast can also make us exaggerate the differences. Heidi Maibom (2019), for example, suggests that the constant contrasting has made us view guilt as overly dependent on autonomously endorsed norms and shame as overly dependent on the judgment of other people (heteronomous norms), and to make us view guilt as overly positive even though it is also associated with major depressive disorder, and shame overly negative and pathological even though it can have a constructive role in rebuilding one's character.

So, even though the focus on shame has brought along many valuable comparisons of guilt and shame, there is also a reason to be vigilant about the fact that such comparisons may tempt us to exaggerate both the similarities (the assumption that they belong to the same class of emotional phenomena) and also the differences. Thus, philosophical research would also benefit from analyses where guilt would be a more independent object of study. This is precisely what this dissertation is aimed at.

These were all remarks about the philosophy of emotion during the recent decades. What about attention to the emotion of guilt in the narrower bounds of ethics? After all, guilt is one of the most paradigmatic moral emotions—the emotion that is supposed to map with moral wrongdoing. So, surely moral philosophers must have a lot to say about it? Let us first consider normative ethics, which deals with primary-level questions concerning which actions are right and wrong and which character traits are good and bad. Those who are interested in emotions are usually virtue ethicists who focus on the role of emotion in virtuous character. Yet guilt does not receive treatment in the classic works of Aristotle (whereas shame did). So, there is no tradition of talking about the virtue of guilt. And, further, as with shame, there is also a sense in which a truly virtuous sage should not have much reason to feel guilt—because they already act virtuously, there should not be much occasion for guilt. At least based on the traditional understanding of guilt as linked to moral wrongdoing, the virtue of guilt could never stand on par with the virtue of justice or that of charity. At best, it could be a secondary virtue for the morally imperfect—something that might be important for the development of virtue, but not as important for the already virtuous sage. Or else, if we put the focus not on guilt as such but on the fact that it is a manifestation of a general sensitivity to right and wrong—John Deigh, for example, associates guilt with our general sense of justice (Deigh 1982), which is basically just a synonym for general moral sensitivity (Greenspan, 1995, p. 111)—, then it becomes part of a ubiquitous metavirtue similar to integrity that may seem redundant precisely due to its ubiquitousness. If one already possesses all the virtues, then they should already have such general sensitivity, and there may seem to be no further need to talk about a separate virtue of moral sensitivity. So, even though virtue ethicists should be especially interested in emotion, guilt in particular has not ended up being the emotion of choice. But what about those who are more focused on deontic rather than virtue ethics? After all, deontic ethics is precisely the field that deals with what guilt is usually concerned with—the moral wrongness of actions. But, of course, deontic ethics is not particularly

hospitable to guilt because it is concerned with deeds rather than emotions. So, eventually, barely anyone ends up talking about guilt. Even though it is a topic that is on everyone's land. It is the bridge that, as also put by Patricia Greenspan, links the ethics of virtue to that of duty (1994, p. 57). It links the sphere of deontic ought with emotions and character.

Continuing with the interest in guilt in ethics, let us secondly consider meta-ethics—the part of ethics that deals with secondary-level questions like what it means to make a moral judgment. To the extent that meta-ethics investigates the nature of moral judgment, guilt can become important within the sentimentalist tradition that seeks to explain our value judgments in terms of emotions—not just as response-independent beliefs but as states that are constitutively dependent on a certain type of emotional response. Can judgments of moral wrongfulness be explained in terms of some particular emotion type(s)? As noted by Justin D'Arms and Daniel Jacobson (2022), it has usually been linked with anger, not so much with guilt. I suspect that one of the reasons behind this choice may be that self-reflective emotions are often seen to build on other-directed emotions—so guilt is often modelled on anger, as a kind of anger or blame turned towards the self (M. C. Nussbaum, 2016, p. 128ff). If this is true, then anger seems like an emotion of more primary importance, and it should not come as a surprise that it has traditionally received more attention. The tides may be turning, though. As D'Arms and Jacobson (2022) note, anger is too coarse-grained to map exclusively with moral issues. It is not just a response to moral offences but goal frustration in general. And its more fine-grained versions, like second-personal resentment and third-personal indignation, are arguably just cognitive sharpenings of anger that involve a thought that invokes a moral complaint (D'Arms & Jacobson, 2003). For those reasons, D'Arms and Jacobson (2022) have recently suggested that guilt might actually be better suited to explain moral wrongfulness, especially if we construe it on the basis of the motivational theory of emotion that does not already assume, like the cognitivist one, that guilt must involve some response-independent moral judgment. But these are just relatively recent developments.

One of the more general problems with regard to the treatment of guilt within moral philosophy, however, is the fact that, given the interests of the discipline as a whole, moral philosophers are likely to be interested in guilt mainly to the extent that it is indeed linked to moral judgment. So, if we are interested in cases without wrongdoing, then, even if those are to be regarded as guilt, these may not be the kind of guilt that moral philosophers are particularly interested in. So, even though their expertise on moral judgment may be of vital importance for distinguishing the cases that truly involve no sense of moral wrongdoing from those that nevertheless do so, they might themselves not be particularly interested in the kind of guilt that is not neatly linked to moral wrongdoing.

Last but not least, I suspect that one of the reasons why guilt has received so little attention might also have to do with the partly already mentioned fact that guilt is on everyone's land. Doing research on guilt may feel like placing yourself at a rather dangerous crossroads of all too many different fields and debates. It requires expertise or at least some knowledge in normative ethics, meta-ethics,

emotion theory, social psychology, and perhaps even clinical psychology if you are interested in especially hard cases like survivor guilt. Of course, there is a sense in which other emotions are objects of interdisciplinary research as well, yet the problem seems to compound in the case of guilt because, as partly already mentioned, guilt seems like an oddball among other emotions. First, guilt seems especially closely linked to morality and thus requires additional expertise in both normative ethics and meta-ethics—an expertise that we do not need to discuss the evaluative profile of fear, for example. Secondly, compared to more basic emotions like fear or anger, the status of guilt in our classifications of emotions is unclear because many have suggested that we should not count it among the basic emotions (Elison, 2005; Griffiths, 1997, p. 100; Kurth, forthcoming; Maley et al., 2019; Ortony, 1987). But how should we categorise it then—a higher cognitive emotion, not an emotion at all, an identificatory mechanism? Furthermore, some of the best philosophical work on guilt has been written through the lens of cognitivism (Deigh, Greenspan, also Zhao as a more recent author), yet cognitivism has fallen decidedly out of favour as a general theory of emotion. It is not at all clear how to navigate all the peculiarities of guilt and guilt research.

As can be seen, there are many reasons why the emotional phenomenon of guilt has not received much attention. There are relatively few careful analyses of guilt in its own right, and the number of people who would have dedicated more than one article to guilt is relatively low. There are of course some noteworthy exceptions: Herbert Morris (1976, 1987, 1999), John Deigh (1982, 1999, 2019), Patria Greenspan (1992, 1994, 1995), more recently also Andrew Ingram (2017, 2020) and Michael Zhao (MacKenzie & Zhao, 2023; Zhao, 2020). There are also some important one-paper/chapter contributions: Gabriele Taylor (1985), David Velleman (2003), Jesse Prinz and Shaun Nichols (2010), Jeffrey Murphy (2012), Corey Maley and Gilbert Harman (2019), Justin D’Arms and Daniel Jacobson (2022). But overall, conceptual work on guilt in its own right is scarce.<sup>6</sup>

### 1.2.2. Why do we need a philosophical account of guilt at all?

Of course, it might be pointed out that guilt and cases of guilt without perceived wrongdoing have nevertheless received attention from the other sciences—as already mentioned, social psychology has been at the forefront here. Is that not sufficient? Prominent social psychologists like Roy Baumeister, with his colleagues, have already directed us towards an alternative route that we might take in order to unify guilt. They suggest that instead of viewing guilt as a matter of intrapsychic evaluation of the self as morally blameworthy, we should instead view it as an interpersonal matter of causing harm to other people (Baumeister et al., 1994). And reading their narrative studies (Baumeister et al., 1995a, 1995b; Baumeister & Wotman, 1994) is likely to provide a rich resource for someone

---

<sup>6</sup> I am only concerned with conceptual work on guilt within the analytic tradition. As such, it is unfortunately beyond the scope of this dissertation to elaborate on discussions of guilt in Nietzsche scholarship and phenomenological treatments of existential guilt.

who seeks a deeper understanding of guilt for harm to others. The idea of guilt as a matter of empathising with the harm that has befallen others is also prominent in the works of Martin Hoffman (2000), and continues to live on, even though as just one form of guilt,<sup>7</sup> in Francesco Mancini's and Angela Gangemi's (2021) account of altruistic guilt.<sup>8</sup> So, is there really a further need for philosophical exploration, a need other than bringing philosophy more up to date with other sciences?

I believe there is. In particular, I am concerned that the focus on harm to others has obscured the role that morality has to play in our feelings of guilt. Perhaps the interpersonal view is not entirely incompatible with giving a role to morality, but for the most part, those connections seem to fade. When Baumeister suggests that instead of moral norm violations we should shift our attention to harm to others, the keyword of morality is already left out of the picture. As such, guilt loses its standing as a moral emotion in the traditional sense. I will have more to say about the different senses of being a moral emotion in chapter 3, but what I mean by moral emotion here is an emotion, the explanation of which involves a moral concept.<sup>9</sup> The traditional description of guilt as a painful feeling about having done something morally wrong fits the requirements of being a moral emotion perfectly because it involves a reference to moral wrongfulness. Mere reference to harm to others, however, loses the tight connection with morality as there is no reference to a moral concept anymore. I think this is like throwing the baby away with the bath water: even though such guilt feelings lose their connection with moral wrongness, they are nevertheless strongly connected to moral badness. Even though people who feel survivor guilt need not think of themselves as having done something wrong, they are nevertheless deeply concerned with the fact that it is comparatively unfair that they came out of the catastrophe alive while others who were no less worthy than them did not. Similarly, those who

---

<sup>7</sup> But nevertheless the one that is supposed to explain the cases of mere harm to others and survivor guilt.

<sup>8</sup> For those who may be wondering why I am not referring to the research done by Tangney and the TOSCA measure: even though her research provoked a great deal of controversy about shame, it did not add much to our knowledge of guilt. A growing number of authors (Kurth, forthcoming; Maibom, 2019; Maley, 2015; Maley et al., 2019) have criticised her work and the TOSCA measure of assuming what yet needs to be proved—she claims it as a result of her research that guilt has pro-social action tendencies (whereas shame does not), but the problem is that she operationalised guilt in terms of pro-social action tendencies to begin with, so the “result” is already built into the assumptions. For these reasons I have decided not to use those studies as a starting point to guide my dissertation and have taken a step back to the previous and more prominent work of Baumeister, which has stood the test of time, and remains an influential reference point.

<sup>9</sup> This traditional understanding of moral emotions comes from John Rawls, according to whom moral attitudes are distinguished from natural ones by the fact that “the person’s explanation of his experience invokes a moral concept and its associated principles. His account of his feeling makes reference to an acknowledged right or wrong” (Rawls, 2005, p. 481). I will have more to say about different senses of the term ‘moral emotion’ later on (section 3.1).

have been the unfortunate causal agents of other people's deaths view themselves as contributing to a morally bad state of affairs—that an innocent person who in no way deserved such a fate was killed. In both cases, the sentiment reflects a deep concern with moral matters like justice and desert. I will soon have more to say about my position, but let me first address why some might be sceptical towards any kind of prospect that guilt should be considered as a moral emotion in a traditional sense.

Namely, it might be pointed out that philosophers have long had the tendency to overmoralise the shapes of emotion types (D'Arms & Jacobson, 2000).<sup>10</sup> For example, anger has been interpreted as a response to injustice, whereas it can actually be a response to goal frustration in general (imagine being angry at a tree that has fallen on the road so that you cannot drive past it). Similarly, envy has been viewed as a response to perceiving the other person as unfairly having more than the subject, though such unfairness need not always be a part of the emotion. It seems that most emotions do not have a moral shape. Perhaps we have traditionally overmoralised the shape of guilt as well, and shifting attention from moral norm violations to harm to others is only to be welcomed. In fact, Jesse Prinz has argued that no emotion, guilt included, is a moral emotion in a strict sense (2009). If this is the case, then it might be suggested that my worries about the non-moralisation of guilt are groundless, and there will be no point in trying to fix something that does not need repair.

So, it might be said that no one should overmoralise the shape of guilt, and, given my background in moral philosophy, I am especially biased toward overmoralising the shapes of emotions. To be fair, the problem is probably also a more general one. It seems that everyone interprets guilt through their own lens, even some of the most prominent authors. Social psychologists like Baumeister speak of guilt as essentially linked to interpersonal relationships. This should not come as a surprise because social psychology is in great part precisely about that—social relations. Martin Hoffman, a clinical and developmental psychologist, whose main topic of research is empathy, tells us that guilt is all about empathy. John Deigh, who is a professor of philosophy and law, gives us an account of guilt that is about a (perhaps somewhat legalistic?) transgression of moral rules. As such, it can hardly come as a surprise that, given my background in moral philosophy, I am tempted to argue that guilt is still a moral emotion. In all those cases, it could be argued that the choice of focus says less about guilt and more about the researcher and the field that they are coming from.

It is probably true that different disciplines, moral philosophy included, approach the emotion of guilt with certain biases. One important way to fight against such biases is to ground one's study in what it is that people actually try to capture when they talk about guilt in their daily lives. This is the reason why I am focused on cases of guilt for merely causal harm and survivor guilt—both

---

<sup>10</sup> It should be noted that D'Arms and Jacobson themselves do not entirely exclude the possibility of truly moral emotions. In fact, they leave it open whether guilt might be one. I am currently just referring to more general phenomenon of overmoralisation that they talk about.

quite well-known real-life cases, cases known from both people's autobiographies and psychological studies. Even though I will be offering a different interpretation of those cases compared to social psychologists, my study is nevertheless grounded in the very same cases rather than mere armchair speculations.

And even though different disciplines have different biases, we should not forget that expertise in a certain field does not only have the potential to make us biased but also gives us a deeper understanding of certain aspects of the phenomena that we study and thus has a potential to enrich the interdisciplinary theory formation on guilt feelings. I do not wish to argue that moral philosophy should have dominion over morality. There is, of course, a lot of important empirical work that has been done on moral judgment as well (for example, Jonathan Haidt's studies on the moral foundations (Haidt, 2012; Haidt & Joseph, 2004)). But there are certain aspects of morality that philosophy in general and moral philosophy in particular are in a special position to disentangle because those are the disciplines that are especially strong at conceptual clarification and understanding the normative structure of moral judgment. So, when arguing over whether guilt remains a moral emotion in a traditional sense, then, regardless of whether we argue for or against the idea, it is hard to see how the question could be settled without expertise in moral philosophy.

Furthermore, the fact that, historically, philosophers have perhaps gone too far with overmoralising the shapes of emotions, should not make us give up the idea that guilt might nevertheless be a moral emotion in a traditional sense. If there are any moral emotions in a traditional sense at all, then guilt would seem like the most obvious candidate. It is like one of the last men standing, and whether it will remain standing once adjusted for a wider range of cases is an interesting question on its own. And perhaps the fact that guilt has been considered somewhat of an oddball among other emotions may actually play in its favour when it comes to retaining its position as a moral emotion.

## **1.3. The account of guilt defended in this dissertation**

### **1.3.1. The main ideas**

So far, I have spoken of my topic—guilt without perceived wrongdoing—more broadly. Now, it is time to unveil the specifics of the position that will be defended in this dissertation. Since I want to be especially clear about the overall contours of my project, this part is laid out in philosophically more rigorous terms.

The quest for making sense of guilt attributions in cases without perceived wrongdoing is essentially what Andrea Scarantino has labelled as the Folk Emotion Project—it aims to offer a descriptive definition of a particular emotion category that can be found from ordinary folk use (Scarantino, 2012, p. 364). As such, I am interested in guilt as a phenomenon that people try to capture in their ordinary talk about guilt; I am interested in guilt as it presents itself in the

narrative flow of human life. This should be distinguished from the so-called Scientific Emotion Project that seeks to uncover the natural, biological kinds that our folk categories of emotion may partially coincide with (Scarantino, 2012, p. 364). More specifically, I will not be concerned with covering all folk use of guilt but only the subset that extends guilt to cases without perceived wrongdoing. I aim to show that this folk discourse is not just a confused way of speaking that might, at best, be based on some loose family resemblance, but reveals a coherent concept of guilt that systematically bleeds out of the narrow boundaries of moral wrongdoing.

The key point I advocate for is that, to understand guilt without perceived wrongdoing, we have to shift our attention from the moral wrongness of actions to the moral badness of states of affairs. I reconstruct the core relational theme of guilt, i.e. that which guilt is about (sometimes also referred to in terms of the formal object),<sup>11</sup> as a matter of contributing to a morally bad state of affairs. According to my account, one can contribute to a morally bad state of affairs in either a causal or constitutive manner, where both are understood in counterfactual terms, meaning that had the subject done something differently, had they received a different fate, had they been at a different place etc., the state of affairs that we are concerned with would be at least somewhat less bad. In cases of causal harm, the subject views themselves as contributing to the causal chain that ended with an innocent person suffering undeserved harm. As I will argue, the harm is not just a non-moral matter, but the subject must take a moralised view of harm and consider it morally bad. The moral badness is also intensified by the fact that the suffering of the victim is viewed as undeserved. In cases of survivor guilt, the subject views the fact that they survived as constitutively contributing to a comparatively unjust state of affairs that consists in the fact that they came out of the catastrophe alive, while others who were equally worthy did not. Comparative injustice, too, is a form of moral badness.

Emphasising the moral badness makes the account stand out in two important respects. First, it contrasts with the traditional idea that guilt is about moral wrongdoing. While the traditional view is focused on the moral wrongfulness of the agent's own deed (thus involving the deontic ought of *what I ought to do*), cases without moral wrongdoing are focused on the moral badness of the relevant state of affairs (the evaluative ought of *what ought to be*). Secondly, the account stands in contrast to the interpersonal approach that views guilt as a matter of simple harm to others and leaves morality out of the picture. Even if guilt's tie to morality does not come from its link with the deontic sense of what we ought to do, it is nevertheless grounded in the evaluative sense of what ought to be. As such, guilt continues to reflect a deep concern for morality and can rightfully be considered a moral emotion in the traditional sense.

---

<sup>11</sup> I will elaborate on the notion of core relational theme at greater length in section 2.4.

Let us now move to how the value-based construal of guilt should be integrated into a general theory of guilt. While I do suggest that viewing oneself as contributing to a morally bad state of affairs is the right way to understand the guilt feelings of those who are prone to feel guilt for cases without wrongdoing, I do not think that this should be adopted as a general account of all guilt. Instead, I will argue that the fact that guilt feelings bleed out of the bounds of deontic morality is a symptom of a value-based<sup>12</sup> moral outlook that views the good as explanatorily prior to the right, so the rightness and wrongness of actions is derived from their relation to the goodness and badness of certain states of affairs. Since people with value-based moral outlook<sup>13</sup> are especially focused on what happens out there in the world, there is a sense in which whether one has contributed to a morally bad state of affairs in a morally culpable way or by no fault of their own is not really the most salient feature of the situation—the most salient feature is the fact that they have nevertheless contributed to a morally bad state of affairs, which explains why the emotion is structurally so similar that it makes sense to call it by the same name—guilt. For people who do not share such value-based outlook, however, the focus is going to be elsewhere. For example, some deontologists view the right itself as a normative primitive,<sup>14</sup> so their focus will be solely on the wrongness of the action, which will keep their guilt strictly within the bounds of the deontic. This can explain why, for some people, it will make no sense at all to talk about guilt without moral wrongdoing and why people can often misunderstand each other when they talk about their guilt feelings.

All this has an important upshot for our theorisation of guilt. We are used to thinking that what may vary about guilt feelings are simply the moral norms for the violation of which people feel guilt. For example, Kantians would feel guilt for benevolent lies while utilitarians might not, religious people would feel guilt for the violations of religious morality, while those with a secular moral outlook would not. All this leaves the basic shape of guilt feelings themselves intact—guilt is still felt about perceived wrongdoing. But as my case studies of guilt without wrongdoing suggest, guilt feelings do not only vary based on the content of moral norms, they can also vary based on the normative structure of one's moral outlook, and this variance can have a dramatic effect on the very shape of guilt

---

<sup>12</sup> Some theorists like Michael Stocker (1973) and Peter Vallentyne (1987) would call such account axiological. The reason why I have chosen to speak in terms of values is because I have found it less ambiguous, as axiological terms can also apply to states of character but my focus in this dissertation is on the kind of goodness that can be attributed to states of affairs.

<sup>13</sup> There are several reasons why I talk about 'outlook' rather than 'theory'. First, a theory should already take a more specific form, and be related to specific values. It is for the same reason that we do not call consequentialism a theory, while we do call hedonistic utilitarianism a theory. Consequentialism just characterises the general form that is characteristic to many different theories. Secondly, since I am talking about how people in general perceive the matters of morality, I do not think that we must necessarily attribute anything as ambitious as a coherent theory to people.

<sup>14</sup> I take this term from (Hammerton, 2025, p. 323).

feelings. For those who have a more traditionally deontological moral outlook, guilt will be first and foremost felt about moral wrongdoing as such, while for those with a value-based moral outlook, guilt will be felt first and foremost about contributing to a morally bad state of affairs.

Last but not least, let me be clear about what will and what will not be achieved by my analysis of guilt as a matter of contributing to a morally bad state of affairs on the front of the appropriateness of guilt, so that my project would not raise any false expectations. It is beyond the scope of this dissertation to defend the idea that cases of merely causal guilt and survivor guilt are objectively appropriate instances of guilt. As already noted, I doubt that my account of guilt as a matter of contributing to a morally bad state of affairs will resonate with those who have a more narrowly deontological moral outlook, and who do not derive the right from the good. The dissertation can help them better understand those who report guilt without wrongdoing, so that they will be less inclined to attribute mistaken beliefs about moral wrongdoing to them. But one may still find that guilt without perceived moral wrongdoing arises from the wrong way of approaching morality. This would reflect a more fundamental disagreement about the sources of normativity, which is a debate well beyond the scope of this dissertation. Furthermore, even among those who do consider the good as prior to the right, there may be substantial disagreements over what counts as good—which values are of fundamental importance. For a hedonistic utilitarian, who considers pleasure as the only thing valuable in itself, survivor guilt over the sheer fact of comparative injustice may strike as misplaced. Settling debates over whether certain forms of justice, like comparative justice or desert, have intrinsic value or not would take us deep into value theory, and such inquiry will not be forthcoming either. So, even though I will be clarifying what kind of values are at stake in different instances of guilt, I do not have the pretension of arguing that this is the one and only way to approach value. This is the sense in which my project is of a more descriptive rather than normative kind—I simply try to capture how people may intelligibly construe cases of guilt without moral wrongdoing.

However, we do not usually think that the traditional account of guilt in terms of moral wrongdoing would be a bad account of guilt unless it provides us with a full moral theory of right and wrong. In the same vein, I do not think that it should be considered a failure if the value-based account of guilt does not provide a complete theory of value (or settle the debate concerning whether the good should indeed be considered as prior to the right or not). There is, nevertheless, an important goal that it does achieve: it goes to show that guilt without moral wrongdoing can be *subjectively fitting*. An emotion is subjectively fitting if the subject themselves perceives their emotion as representationally accurate.<sup>15</sup> So, regardless of whether it can indeed be objectively said that the subject has contributed to a morally bad state of affairs (or whether our sense of morality should be focused on the goodness of states of affairs to begin with), the subject

---

<sup>15</sup> For a more complicated picture, see 2.4.

themselves would tend to agree that they have (and that what happens out there in the world is morally important). This means that the experience of guilt without perceived wrongdoing is nothing like the experience of recalcitrant emotion, like fear of harmless snakes or spiders. The latter are cases where the subject is at odds with themselves: on the one hand, they know that there is nothing to be afraid of, yet on the other hand, they are terrified. Cases of guilt without moral wrongdoing need not involve any such inner tension. My analysis shows that merely causal guilt and survivor guilt can be subjectively fitting emotional experiences, because the subject is likely to construe the situation in terms of contributing to a morally bad state of affairs. Thus, rather than being blatantly irrational instances of guilt, merely causal guilt and survivor guilt can form a perfectly meaningful part of one's life and reflect a sincere and deep concern for morality.

### 1.3.2. Methodology

My project could be viewed as a mixture of Carnapian explication and the method of reflective equilibrium (Brun, 2020). On the one hand, I am concerned with explication—I want to capture the wide folk concept of guilt, which includes cases without perceived wrongdoing, into a neat definition that could be used in philosophical and psychological discussions of guilt. This kind of approach should be contrasted with prototype-centred approaches that seek to identify the most prototypical case of guilt, and then suggest that all else can be labelled as guilt based on its similarity to the prototype. I have no qualms with the prototype-based approach. I am interested in guilt, as it is understood in a certain folk psychological discourse, and since there will always be some variance in how people apply the word guilt, it is unrealistic to expect that one definition could ever capture all the folk usage. In that sense, the prototype approach—with its focus on prototypical cases and the idea that non-prototypical cases belong to the same sphere in virtue of some family resemblance with the prototypical cases—is a perfectly reasonable approach (Scarantino, 2012, p. 365). Neither do I have any qualms with the idea that the most prototypical cases of guilt do indeed involve perceived wrongdoing. My aim is simply different because I am interested in coming up with a concept that would cover as much of the common use as possible without losing theoretical fruitfulness in the sense that we could still make generalisations about all those cases *in corpore*. So, instead of pinpointing a prototype, I want to throw a lasso on the field of guilt and capture as many instances as possible. On the other hand, as argued by Brun (2020), explication must be complemented by seeking a reflective equilibrium. We are not just dealing with a one-directional task of looking at a fixed set of cases, and coming up with a perfect definition, but with a more dynamic challenge that requires finding a balance between our theoretical conceptualisation of guilt and the cases that it covers. In other words, the cases will not only inform our

conceptualisations, but the other way around as well—our conceptualisations will also further shape and clarify the contours of the cases.<sup>16</sup>

In my choice of methodology, I have also been inspired by Kristjan Kristjánsson's method of critical conceptual revision (Kristjánsson, 2006, p. 11), which too tries to find a balance between theoretical characterisation and folk use, in a way that our theories would retain the truth of the greatest number of folk conceptions. Indeed, Kristjánsson is explicit in his reliance on the reflective equilibrium (p. 9), but I think that his ambition to retain the truth of the greatest number of folk conceptions is also reflective of Carnapian aspirations. As already explained, those two methods are best seen as complementing each other.

To the extent that I am trying to capture the folk concept of guilt, my inquiry is of a more descriptive rather than strongly normative kind. I do not aim at radical conceptual re-engineering. But, of course, as I aim at explication, and there may be more than one way to capture the greatest number of cases, it cannot be said that my inquiry is fully neutral and value-free. I am especially interested in the first-person experience of guilt, especially in the context of the narrative flow of human life. This is basically the same approach that Robert Roberts takes towards emotions in general (Roberts, 2003, p. 57), and should be contrasted with disciplines that look at emotions from a third-person perspective, for example, those that view them as a physiological or neuroscientific phenomena. But interest in emotions in the context of the narrative flow of human life is definitely not exclusive to philosophy or moral philosophy, it also reflects a rather general type of human interest (Roberts, 2003, p. 56) which seems to have been shared by some prominent social psychologists as well—especially Roy Baumeister and his colleagues who have used a narrative method in their study of guilt (Baumeister et al., 1995b, 1995a).

This thesis aims to contribute to theory formation on guilt by suggesting a value-based way to bring unity to guilt feelings without perceived wrongdoing. To the extent that it is possible, I have been committed to writing in an empirically informed manner—relying on both psychological research and also narrative accounts that can be found from less formal avenues such as media interviews, blog posts and autobiographies. Those latter sources are important as well because they help us put emotions into a rich narrative flow of human life.<sup>17</sup> The very starting point of this dissertation lies in the fact that there seems to be plenty of evidence out there that shows that guilt is not just a matter of perceived moral wrongdoing. Relying on the existing psychological research on guilt, I will show that appealing to mere harm to others, as the widely influential interpersonal account of guilt has done, is not the only and also not the best way to explain

---

<sup>16</sup> I do not think that combining the methods of Carnapian explication and reflective equilibrium is in itself new. I am optimistic that most of the work that was formerly known under the label of 'conceptual analysis' more or less reflectively strived at the same goal. But I do think that Brun has been the most successful in capturing the true spirit of this endeavour.

<sup>17</sup> In the same vein, Robert Roberts also justifies the inclusion of psychoanalytic and literary sources for the study of emotion (Roberts, 2003, p. 4).

what is going on in cases of guilt without wrongdoing. The alternative that I offer would enable us to offer a neater explanation to a wider range of cases and would also allow us to see that, in those cases, too, guilt can be a fitting and meaningful part of people's moral lives. However, it is important to note that even though my alternative explanation is in accordance with the existing evidence (in the sense that it does not contradict it, and, at times, accords with the existing evidence even better than the interpersonal account), it is important to note that it does involve a hypothetical element. I am just a philosopher who seeks to enrich the debate by bringing in a more informed perspective on the nuances of moral judgment. Since I am going against the common way how guilt is described in social psychological research, I cannot cite studies that would conclusively prove that my account is better because no such studies have been conducted that would tease out the differences between the interpersonal account and the view that I am proposing. As such, some of my claims will have to be speculative and conclusive proof is not what the reader should expect when reading this dissertation. My aim is simply to offer a plausible and theoretically elegant alternative that would be in line with the existing evidence.

### 1.3.3. Chapter Outline

Chapter 2 lays the groundwork for the rest of the dissertation: it takes a closer look at the problem that drives the whole dissertation—the clash between the standard account of guilt in terms of perceived moral wrongdoing and the counter-cases of merely causal harm and unfair benefits. First, I argue that there is indeed a serious conflict between the two. I will give an overview of both in their strongest form, the standard account and the counterexamples. Then, I will review the main strategies that the standard approach has left at its disposal to respond to cases without wrongdoing—the attribution of irrational judgments (or thoughts) to the subject and denying their emotion the status of guilt—and explain why both are problematic. In response to the strategy of denying those cases the status of guilt I will also take up the question of whether guilt is really an apt label. As already explained, I am more concerned with unity rather than labels, but I will nevertheless review some of the meta-theoretical issues at stake when choosing between different labelling practices, and highlight the advantages of extending 'guilt' to cases without perceived wrongdoing. Secondly, I will further clarify the nature of the clash by showing why the tension will remain even if we gave up on the cognitivist element of the standard theory—because the main problem does lie in the standard account's cognitivist form but the more general idea that the core relational theme of guilt should be viewed in terms of moral wrongdoing. This part of the text will provide a rationale for why I have chosen to organise the discussion in the following chapters by the nuances and contours of the core relational theme rather than emotion theory, and it will also provide the proponents of different theories of emotion the guide to translate all the following discussions into the language of their preferred theory of emotion.

Having established that the most important question at stake with regard to the counterexamples is the core relational theme, I will next move on to consider how we might revise the core relational theme in order to make room for cases without perceived wrongdoing. I will proceed by visiting the attempts to revise the core relational theme in a case-by-case manner. Chapter 3 focuses on attempts to reconstruct the core relational theme of guilt based on cases of merely causal harm to others, and Chapter 4 focuses on attempts to revise the core relational theme based on survivor guilt and other forms of guilt for unfair benefits. Both chapters will proceed in a similar fashion. My main question is: if the relevant type is not to be characterised in terms of moral wrongdoing, then what would be the correct way to characterise it—what is it about the situation that people find so important when they label it guilt?

Chapter 3 focuses on how we should interpret cases of merely causal harm to others—what exactly is it about those situations that stands out to people and makes them label those cases as instances of guilt? The most obvious answer is offered by the interpersonal account of guilt, which was in great part designed to do justice to those cases. According to the interpersonal account of guilt, the core relational theme of guilt should be understood in terms of harm to others (Baumeister et al., 1994). First, I will give an overview of the interpersonal account and explain why it may seem like an especially attractive approach when we view it on the background of critical concerns about overmoralising the shapes of emotion that have been raised by D’Arms and Jacobson (2000) and Jesse Prinz (2009). Even though sympathetic towards the general concerns about the overmoralisation of emotion, I will nevertheless go on to argue that guilt might be an exceptional case that might rightfully deserve to be called a moral emotion in the traditional sense. I will start by pointing out that one of the problems of framing guilt as a matter of mere harm to others is that it leaves no conceptual space for instances where one has caused harm to others and feels sincerely sorry, but the feeling does not quite amount to guilt. Then, I will go on to clarify what differentiates cases of guilt from merely feeling sorry for the harm that one has caused. I will argue that even though guilt need not involve a reference to deontic wrongness, it does nevertheless involve a reference to moral badness, which is built into the cases of guilt for harm to others in many different layers. First, I will suggest that it is plausible that harm to others is going to elicit guilt only if one takes a harm-based view of morality and considers their deontic duties in terms of (the relevant type of) harm as well. Furthermore, I will argue that in all the cases of merely causal harm that the interpersonal theory usually refers to, there is also a second, often unnoticed, layer of moral badness that comes from the fact that harm is seen as undeserved.

Chapter 4 will focus on attempts to revise the core relational theme of guilt in order to do justice to cases of survivor guilt and other forms of guilt for unfair benefits. I will proceed by reviewing the explanations offered by the interpersonal account and David Velleman, and gradually build towards a more plausible account. If my main problem with attempts to explain cases of merely causal harm was that we are not given a sufficiently full account (as the interpersonal

theory leaves out morality), then here my main problem is that theorists often introduce too many unnecessary elements to survivor guilt. First, I will consider the solution offered by the interpersonal account because one of the original pretensions of the interpersonal account was to accommodate those cases as well. The interpersonal account of guilt explains all kinds of guilt via their resemblance to guilt for harm to others, and in the case of survivor guilt this similarity is supposed to lie in the fact that in both cases, the subject benefits, at least seemingly, at the expense of other people (Baumeister et al., 1994, p. 247). By appealing to counter-examples, I will argue that perceiving oneself as benefitting at the expense of others is not a necessary element of all survivor guilt. Secondly, I will discuss the view proposed by David Velleman (2003), according to whom survivor guilt is similar to regular guilt because both involve anxiety about having warranted resentment, where resentment is understood as an emotion “embittered by a sense of injustice” (p. 247)—while regular guilt involves anxiety about having warranted angry resentment, survivor guilt involves anxiety about having warranted envious resentment. However, since he approaches guilt through the framework of the Freudian theory, he puts a great emphasis on the idea that guilt is to be explained in terms of fear or anxiety. I will argue against the idea that anxiety is a necessary element of guilt but I do agree with Velleman’s idea that if there was anxiety about resentment, this would have to be anxiety about warranted resentment. Velleman explains it in terms of normative vulnerability—the idea that one would have nothing to say in defence of his position. As I will argue, once we peel off the layers of anxiety, what this ultimately reduces to is simply the fact that one is an integral part of the comparatively unjust state of affairs.

In chapter 5, I will try to extrapolate a more general theory of guilt from the previously discussed cases. As I will argue, both types of guilt arise from a similar kind of value-based approach to morality where the focus is on morally bad states of affairs. Only in one case, the subject contributes to the morally bad state of affairs in a causal manner (by being part of the causal chain that led up to the morally bad state of affairs) and in the other, in a constitutive manner (by being an integral part of the comparative injustice). Both types of contribution are nevertheless similar, though, because both can be explained in counterfactual terms. Had the subject not driven that road at that time, the accident would not have happened. Had the subject died alongside other people, the result would have been comparatively more just. This emphasises the fact that the subject makes a difference in the world. Both forms of guilt without perceived wrongdoing track this difference. I will also show how this approach is a kind of midway between the traditional account of guilt and the interpersonal account: it combines the outward focus on the interpersonal account, with the idea of guilt as a moral emotion embedded in the traditional account (even though the moral element relates to the bad rather than the wrong). However, I will argue that guilt understood in these terms does not exhaust the whole sphere of guilt. Rather, it just reflects the kind of guilt that arises from a value-based moral outlook that focuses on states of affairs and views the good as prior to the right. Since the focus is on the good, it is only natural that the guilt feelings bleed out from the bounds of deontic morality.

However, for someone who views the right as prior to the good, guilt feelings will remain within the bounds of the. As such, I will end up defending the idea that the very shape of guilt feelings is influenced by one's normative outlook.

There is an important remark I would like to add about the overall logic of this dissertation. As I will first be providing a treatment of guilt for merely causal harm (Chapter 3) and survivor guilt (Chapter 4), and only then present a more general theory of guilt on the basis of those treatments (Chapter 5), this may leave the impression that this dissertation proceeds in a linear manner. To some extent, it does: the account that I present is indeed, in great part, informed by my case studies. But there is also a sense in which the linear structure is just a playful appearance that I have adopted because of the limitations of the medium, structural elegance, and, not least, to avoid burdening the reader with all the back-and-forth of finding a reflective equilibrium that has taken place behind the scenes. Since the thought process behind this dissertation is that of reflective equilibrium, my analysis of cases should not only be viewed as supporting the final account of guilt that I offer, but it is also the other way around. The final account that I offer should be seen as providing additional support to the analyses of merely causal harm and survivor guilt. The reason why I capture guilt in the terms that I do in my case analyses is already partly informed by the account to which I will reach. Neither the ideas in this dissertation nor the text itself have been developed in a linear fashion—all the chapters that follow should be viewed as forming a mutually supportive whole. This should not be seen as a problematic kind of circularity but a natural consequence of relying on the method of reflective equilibrium and having a holistic approach to theory development.

This nonlinear structure also has some important implications on how to read this text, especially for those who are interested in only one of the two problem cases at the heart of this dissertation, either guilt for merely causal harm or survivor guilt. It will not be sufficient to only read the chapter on merely causal harm or survivor guilt. To get a complete account of what I have to say about each type of guilt, I would suggest proceeding in the following order. It would be useful to start from 2.2. where I introduce the relevant type of guilt through numerous examples and take special care to present those examples in their purified form, differentiated from similar instances that can be explained by the traditional account. Next, one can proceed to either Chapter 3 or 4, depending on which special case they are interested. And finally, I would suggest reading 5.1, which explains how those cases fit into my more general account of guilt.

## **2. GUILT AND MORAL WRONGDOING**

In this chapter, I will take a more detailed look at the clash that drives this whole dissertation—the conflict between the traditional analysis of guilt as a painful feeling about perceived wrongdoing and the problem cases of merely causal guilt and survivor guilt. First, I will give an overview of the traditional account in its best form. Secondly, I will give an overview of the problem cases in their best form. I will argue that on the background of other commonly mentioned problem cases like collective guilt and self-disciplinary guilt, the cases of merely causal harm and survivor guilt pose an especially serious problem for the traditional account of guilt. This will serve to explain why I have chosen those cases as the central focus of this dissertation. In my overview of the problem cases I will already try to anticipate some of the ways how the proponents of the traditional account may try to explain those cases away as rather traditional instances, so I will take care to present the problem cases in their purified form where it does not seem particularly likely that one should have irrational beliefs or thoughts of having done something morally wrong. In the third section, I will look at the strategies that are left to the traditional account to deal with the purified forms of those cases—attributing unconscious irrational beliefs to the subject or denying their emotion the status of guilt. I will explain why the former is uncharitable and the latter neglectful. Even though this dissertation is first and foremost about relationships between different emotional phenomena, it is in relation to denying cases without perceived wrongdoing the label of guilt that I will also elaborate on the meta-theoretical issue of how to decide which cases a good theory of guilt should cover and illuminate some of the benefits on extending guilt beyond cases with perceived wrongdoing. Fourth, I will consider how to move forward from the failure of the traditional analysis. In particular, I will argue that changing the theory of emotion will prove to be an insufficient solution, and the more serious debate concerns the core relational theme of guilt.

### **2.1. The traditional account of guilt**

According to a classic cognitivist analysis, guilt is a painful emotion that builds on the judgment that one has done something morally wrong. As such, it consists of two components: its cognitivist form (it requires a judgment or something close to judgment) and moral content (the judgment must be about having done something morally wrong). Next, I will touch upon both of these building blocks. The aim is to explicate the standard theory in its strongest form.

#### **2.1.1. The cognitivist aspect**

I will begin with the cognitivist aspect. It has indeed been very common to associate guilt with certain thoughts or judgments that have propositional content. As put by Herbert Morris (1976): “In guilt the “voice of conscience” speaks and we

formulate in words what is to be and not to be done, words that are spoken and heard” (p. 62), and later echoed in Bernard Williams (1993): “guilt is rooted in hearing, the sound in oneself of the voice of judgement;<sup>35</sup> it is the moral sentiment of the word” (p. 89).<sup>18</sup> This accords well with the cognitive theory of emotion, which identifies the defining element of the emotion as some evaluative cognition, i.e. some cognitive state with evaluative propositional content. There are disagreements about which cognitive state exactly it is and some variance as to how different authors may prefer to label it. Since the problem that I want to highlight—the fact that those accounts are ill-equipped to do justice to cases without perceived wrongdoing—is of a more general kind, all the intricate difference between different forms of cognitivism need not concern us here. So, most of the time, I will simply be referring to the relevant cognitive state by talking about how the subject ‘perceives’ the situation, preferring the word ‘perceives’ precisely for its vagueness. Yet it is worth highlighting that if we are to interpret the traditional account of guilt in the strongest form, then requiring that guilt must involve a full-blown evaluative judgment (M. C. Nussbaum, 2003; Solomon, 2003) is too strong, and we should be open to Patricia Greenspan’s quasijudgmentalist<sup>19</sup> idea that guilt might just involve evaluative thought that the subject need not assent to (Greenspan, 1995), for the advantages of this view, see also (Clarke, 2016). So, even though the subject may judge oneself to be morally blameless, they nevertheless think of themselves as morally blameworthy, i.e. *as if* they were morally responsible (Greenspan, 1992).

Since I want to take the most charitable interpretation as possible, I would also like to emphasise that I do not consider it an essential part of cognitivism that it must reduce the emotion to its cognitive component and deny the existence of an affective element, i.e. a certain phenomenal feel.<sup>20</sup> But the affective component is usually described with less precision—as simply a painful feeling that involves discomfort, and is not considered to be fit to individuate guilt as an emotion type. In fact, some theorists even hold that the feeling component, even though necessary, is subject to variation, so that there is no common phenomenological feel to all instances of guilt (Maley et al., 2019).<sup>21</sup> But even if there was, it would probably be somewhat harder to describe, and thus the burden of individuating

---

<sup>18</sup> See also (Taylor, 1985, p. 85f), who speaks about how the demands of authority become the voice of conscience.

<sup>19</sup> She herself calls her account ‘nonjudgmentalist’ analysis of guilt (Greenspan, 1992), because she wants to distance her account from a requiring a full-blown evaluative judgment, but given that she nevertheless continues to work within the frame of a cognitivist analysis, the account has also been called ‘quasijudgmentalism’ (D’Arms & Jacobson, 2003), which is a label I tend to find more apt as well.

<sup>20</sup> For the sake of simplicity, I am ignoring here the possibility that cognitive states may have a phenomenology as well (Kriegel, 2015a, 2015b), and refer to the phenomenal feel only in reference to the more bodily aspect of emotions.

<sup>21</sup> This idea might also receive some support from Francesco Mancini’s distinction between deontological and altruistic guilt, where the latter is viewed as more closely tied to empathetic identification with the victim’s pain, and the former with the reverence for higher order.

guilt as a unique emotion type, is usually put on the cognitive component. So, even though the cognitive component need not be considered as sufficient for guilt, it is nevertheless considered a key element in the sense that it is on the basis of this element that we individuate guilt as a unique emotion type.

Neither do I think that a cognitivist about guilt would necessarily have to endorse cognitivism as a general theory of emotion. During the last decades, cognitivism has lost its popularity as a general theory of emotion, but that does not mean that it cannot provide an account of how to understand some emotional phenomena. Following Paul Griffiths' (1997) division between affect programs and higher cognitive emotions, it may well be argued that guilt belongs to the category of higher cognitive emotions (see, for example, Elison, 2005).<sup>22</sup> Differently from more basic emotions like fear that we also share with other animals and infants, guilt definitely looks like a more complex emotion. As also highlighted by Robert Solomon:

What makes guilt particularly interesting [---] is the fact that one can *feel guilty* despite the fact that one has done nothing wrong. [---] One might suppose that this feeling of guilt without actual guilt is very much like our earlier example of irrational fear, where one is afraid of something that one knows not to be dangerous. But I think that the cases are quite different, first because no one has proposed anything remotely like a brain mechanism for guilt, as has been so elegantly demonstrated in the case of fear reactions, second because there is a complexity to guilt that far outstrips anything that we saw in primitive fear. (Solomon, 2008, p. 92)

It seems fairly plausible to say that guilt does not have the kind of evolutionarily hard-wired automaticity built into it as there is for more simple emotions like fear (see also Kurth, forthcoming), even though there might be some debate as to whether such automaticity might be achieved through learning.<sup>23</sup>

### 2.1.2. The moral wrongdoing aspect

Let us now turn to the other building block of the traditional account—its moral content. The relevant cognition is supposed to be about having done something morally wrong, where moral wrongdoing is understood in a broad sense, including omissions as well, and implies that the agent could and should have acted differently. I take moral wrongdoing to be more or less synonymous with or to go hand in hand with having violated an internalised moral norm, being morally culpable, morally responsible, morally at fault, morally guilty and blameworthy.

---

<sup>22</sup> Alternatively, it has also been pointed out by Justin D'Arms and Daniel Jacobson, that some more complex emotional phenomena are best understood as cognitive sharpenings of more simple emotions (2003, pp. 137–138). For instance, they consider resentment as such because it involves the idea of being wronged (2003, p. 143).

<sup>23</sup> What I have in mind here, are the cases where one has been conditioned to experience guilt at the presence of certain cues.

Admittedly, some theorists may make fine-grained distinctions between all those concepts,<sup>24</sup> but I am here speaking in broad brush strokes because, for the purposes of my inquiry, those distinctions are not that relevant. Overall, it is my aim in this dissertation to introduce an altogether different paradigm that is less centred on how we might construe the agent (culpable, at fault, guilty, blameworthy) and more focused on the states of affairs that the agent contributes to. So, this is why I will not elaborate on the potential differences between all the abovementioned terms and will confine myself to speaking of wrongdoing in broad brushstrokes.

Linking guilt with moral wrongdoing has been a standard way to conceptualise guilt among a long array of prominent philosophers. John Rawls emphasises that mature guilt links to the violation of the principles of the right rather than the good (Rawls, 2005, p. 482). R. E. Lamb says that we can “*feel guilty about x only if we believe that we are, in a moral sense, responsible for x*” (Lamb, 1983, p. 335). Walter Sinnott-Armstrong says: “Guilt occurs when one’s bad feelings are caused by one’s belief that one’s act was morally wrong” (Sinnott-Armstrong, 2005, p. 200). Martha Nussbaum says: “Guilt is a negative emotion directed at oneself on the basis of a wrongful act or acts that one thinks one has caused, or at least wished to cause” (M. C. Nussbaum, 2016, p. 128). According to John Deigh: “Guilt [---] is an emotion in which the subject’s mind is troubled by a belief that he is guilty of wrongdoing” (Deigh, 2019, p. 55). Greenspan speaks of a discomfiting thought of oneself as “responsible for a wrong” (Greenspan, 1992, p. 293). Different authors may put it into slightly different words, but the basic idea is the same: guilt is linked to moral wrongdoing.

What exactly does this mean? First, it is important to note that the norms concerned are deontic—they have to do with the rightness and wrongness of actions. Deontic judgments are to be contrasted with axiological judgments that concern the goodness or badness that can be attributed to states of affairs or character traits. So, guilt is not first and foremost about evaluating the consequences of one’s actions as bad or considering one’s moral character as bad (this is

---

<sup>24</sup> For example, the proponents of actual consequentialism, who determine moral rightness based on actual (rather than foreseen, foreseeable, intended, or likely) consequences, might suggest that wrongness may come apart from blameworthiness. For example, if one buys a bus ticket to a runaway, who then gets killed in a freak accident while on the bus, actual consequentialists may say that buying the bus ticket was indeed wrong, but that this is an instance of blameless wrongdoing (Sinnott-Armstrong, 2023, p. 19, not his own position though). If the traditional account of guilt as a matter of wrongdoing were combined with actual consequentialism, then some of the problem cases that I will be discussing would indeed be resolved. But to the extent that even actual consequentialists would tie wrongdoing with agency, their understanding of wrongdoing would still not be broad enough to cover all the problem cases that I am interested in. Survivor guilt involves no agency at all, and even though some of the cases of merely causal harm to others that I will be discussing will involve agency, I am also interested in others where there is no agency—cases where one functions as a mere physical body. So, even though I myself will be using the term ‘wrongdoing’ in a narrower sense than actual consequentialists, as implying blameworthiness, my critique will also apply to accounts of guilt that would tie it with wrongdoing in a broader sense that does not imply blameworthiness.

traditionally considered as the sphere of shame). Even though it does not exclude that one may also make those further judgments, the essence of guilt lies in the fact that it involves a deontic judgment concerning the wrongness of one's actions.

Next, let us look at what the notion of morality adds to those norms. One can speak of morality in either a descriptive or a normative sense (Gert & Gert, 2020). When we speak of morality in a descriptive sense, we are concerned with what counts as moral for particular people or cultures—what are the imperatives by which they live their lives. For example, we can speak of Christian morality, Buddhist morality, individualistic Western morality, perhaps also about the morality of gangsters if they abide by a consistent code of conduct. However, when we speak of morality in a normative sense, we are not just concerned with what different people (however correctly or mistakenly) happen to consider as moral, but we are referring to the true or ideal morality that we think everyone should follow. This sense of morality is evident in sentences like “Which action does morality require of us?” or “Is practice *x* moral?”. Morality, in a normative sense, is the main object of normative ethics. This is the discipline that asks questions about what is truly right or wrong. For the purposes of moral psychology, however, we are not concerned with morality in a normative sense. After all, people feel guilt for what *they* consider as violations of morality, not for what is necessarily a violation of the *objectively correct* morality (if there was any). For example, if a person perceives sex with contraception as morally wrong, then they are bound to feel guilt about having engaged in such sex even if it should actually be the case that there is nothing objectively wrong with it. Morality in a normative sense becomes relevant only when we start assessing the moral appropriateness of guilt.<sup>25</sup> For example, someone with a secular mindset might criticise guilt for sex without contraception as morally inappropriate guilt. This dissertation is not concerned with the moral appropriateness of guilt. Its focus is on illuminating the subjective experience of guilt.

When focusing on moral norms in a descriptive sense,<sup>26</sup> there are two types of definitions: formal and material ones (Wong, 2013). Formal definitions do not make any restrictions on the content of moral norms, they just specify the normative character of moral norms. In philosophy, those formal requirements are traditionally taken to include three features outlined by Hare (1952): prescriptiveness, universalisability and overridingness. Prescriptiveness just refers to the fact that moral norms are action-guiding, they tell us what we ought or ought not to do. According to the criterion of universalisability, a norm counts as a moral norm for the subject if they think that other relevantly similar people in relevantly similar situations should adhere to the norm as well. So, one cannot consider it morally right to lie to another person if they themselves would object to other

---

<sup>25</sup> See more on the importance of differentiating between moral appropriateness and the fittingness of guilt from (D'Arms & Jacobson, 2000).

<sup>26</sup> The following criteria will be necessary for moral norms in a normative sense as well, but not sufficient: their fulfilment does not grant that the moral norms are the objectively correct ones.

relevantly similar people lying in a relevantly similar situation. To better accommodate the norms of personal morality, like a writer's norm to write one page every day, it might be suggested that we loosen the criterion of universalisability, so that rather than requiring that other people must necessarily adhere to the same norm, the subject should at least not object if other similar people in similar situations do so. Overridingness refers to the fact that the norm is considered especially important and serious so that it cannot be overridden by other, more conventional norms. Similar criteria are also reflected in psychological research. In order to differentiate moral norms from merely conventional norms, Elliot Turiel (1983) has suggested that moral norms are more serious and authority-independent (he adds a material criterion as well, but we will get to that). The point about seriousness is similar to the idea that moral norms override other norms, and the point about authority-independence, similarly to the criterion of universalisability, refers to the fact that moral norms are not just based on someone's arbitrary command but are perceived as objective. It refers to a difference that even children understand: a conventional norm to call a teacher by their family name may lose its force if the teacher allows their students to call them by their first name, but even if a teacher says that it is all right to push another child off the swing, that would not make such action permissible (Miyazono & Bortolotti, 2021, p. 107).

Some have been concerned that formal definitions are too broad. For example, Philippa Foot (1969) has pointed out that if some eccentric thought that they ought to clasp and unclasp their hands, this could be counted as a moral norm by the formal criteria. In order to exclude such idiosyncrasies, a further requirement might be added that ties moral norms to what benefits and harms people. Turiel, too, completes the formal criteria of importance and authority-independence by adding that moral norms involve harming, injustice, or infringement on individual rights (1983). A material definition has also been preferred by Jonathan Haidt, who links morality to "the interests and welfare of the society as a whole or at least of persons other than the judge or agent" (Haidt, 2003, p. 853).

When it comes to illuminating the nature of guilt, I do not find it fruitful to place material restrictions on the content of moral norms. Formal definitions have the advantage of greater flexibility. After all, people can feel guilt for all kinds of things. Failures of personal morality, like breaking a diet and overspending, are common objects of guilt. Also, if a Gauguin<sup>27</sup> finds it his personal imperative to pursue his creative ambitions even if it means abandoning everything else (including his family), then this is that Gauguin's personal morality and this is what this Gauguin would feel guilty about if he happened to fail his cause. In that sense, I would suggest that we follow Gabriele Taylor's footsteps, and speak of "personal morality", which is "not thought of just in terms of adhering to or breaking certain moral rules, but is taken to include personal morality, a person's

---

<sup>27</sup> Following in Bernard Williams's footsteps, I am talking about "a Gauguin" as representing a type of artist willing to sacrifice everything else (e.g., leaving his family) for his creative pursuits (B. A. Williams, 1976, p. 117ff).

own view of how he ought to live and what he ought to be” (1985 p. 77, see also Sludds 2009 pp. 69–71). In other words, this includes a person’s own sense of what he, in particular, should do or be, which need not be fully universalisable in the sense that the person does not need to hold other people to the same standards.<sup>28</sup>

As such, a formal definition is sufficient. Moral norms, in the sense that we are concerned when talking about guilt, are those internalised norms that the subject considers as especially important, authority-independent, and universalisable, at least in a weak sense. They concern what the person sincerely feels that they ought or ought not to do. There may, of course, be debates regarding how each criterion should best be formulated, but for the current purposes, this initial gloss should suffice.

To bring the two components of the traditional account together, the overall idea of the traditional account is that a person feels guilt if they hold a judgment or at least have a thought that they did not do what they ought to have done, that they violated a relatively important internalised norm of theirs.

## 2.2. The problem cases

The idea of guilt as an emotion that centres around the judgment or thought that one has done something morally wrong faces the challenge of accommodating a number of potential problem cases. Those counter-examples include guilt for mere harm to others, guilt for unfair benefits (including survivor guilt), self-disciplinary guilt, collective guilt and character guilt.<sup>29</sup> In this section, I will discuss all those cases. First, I will review self-disciplinary guilt, collective guilt and character guilt, and explain why I do not think that they pose a particularly serious problem to the traditional account of guilt. Then, I will proceed with cases of merely causal harm to others and survivor guilt as more serious counterexamples. These are the cases that this dissertation is especially interested in.

---

<sup>28</sup> This resembles formal rather than material definitions of morality in the sense that it is flexible—whatever the agent considers as his duty counts. (See more about material and formal definitions from (Haidt, 2003) and (Wong, 2013).) However, it differs from classic formal definitions because it does not require universality. But the sense of morality is nevertheless often used in everyday language when we speak of different people’s “morality”. For example, we could speak of “Nietzsche’s morality” in the sense of what Nietzsche considered as his ideals of life, or “Gauguin’s morality”. Even though not all people would agree with such moralities, it still makes sense to call them morality because they reflect the particular person’s sense of right and wrong in their life.

<sup>29</sup> This categorisation mirrors the four types of cases that are presented in Maley & Harman 2019: guilt about harm to others, self-directed guilt, religious guilt and survivor’s guilt. But there are also some differences. First and foremost, I have tried to move beyond individual cases and trace those case back to more general categories. Secondly, it does not seem to me that Maley and Harman present harm-to-others cases as problem cases for any of the accounts of guilt that they mention. They simply seem to refer to this type of guilt as one of the main types that a good theory must accommodate. I, however, have put the focus on this subset of harm-to-others cases where there does not seem to be any moral transgression.

### 2.2.1. Collective guilt

I will begin with collective guilt. For example, someone might feel guilt for the war crimes committed by their fellow countrymen. Collective guilt is often posed as a counterexample to the traditional account of guilt because the subject who feels guilt need not have done anything morally wrong (Morris, 1987). This type of guilt has been used to shed doubt on the idea that guilt can be correctly characterised as an emotion of self-assessment, and it has even been suggested that our very inquiry into the nature of guilt feelings should start from looking at both individual and collective guilt (Gilbert, 1997). While I do agree that in the context of global politics, collective guilt feelings are an important matter and that a full account of guilt should also give us an idea of how guilt feelings can extend to the actions of the group to which one belongs, I do not think that our inquiry into guilt should necessarily have to start from collective guilt feelings. In its paradigmatic, non-vicarious form and compared to the paradigmatic, non-vicarious forms of other emotion types, guilt is nevertheless an emotion of self-assessment. Let us compare it with anger. Anger is not paradigmatically an emotion of self-assessment, instead it is focused on the offensive actions of other people. Guilt, however, is paradigmatically focused on our own deeds. Both anger and guilt can take vicarious or collective forms. One can feel guilt on behalf of their group, one can also feel anger on behalf of their group. But since both are equal in the latter respect, we can nevertheless say that, comparatively, guilt is more of an emotion of self-assessment than anger—which is precisely what I think is usually meant when emotion theorists describe guilt as an emotion of self-assessment. Furthermore, it can be argued that to the extent that the subject views their group as an extension of the self, there is a sense in which collective guilt continues to be an emotion of self-assessment as well. But even though we can argue about how exactly we should construe the self of guilt, none of that refutes the idea that the core relational theme or the basic shape of guilt nevertheless concerns viewing one's deed as morally wrong. Even in the case of collective guilt, there is nevertheless someone with whom one identifies who did indeed do something morally wrong. A similar point seems to be emphasised by one of the leading proponents of the traditional view, John Deigh, as well. He draws attention to the fact that guilt feelings felt over wrongs committed by one's nation or community "mimic the feelings that the individuals responsible for the wrongs experience of should experience" (Deigh, 1999, p. 319). In other words, they seem to be derivatives of the non-vicarious forms of guilt feelings. In my inquiry, I am concerned with a more narrowly guilt-specific question that pertains to the core relational theme of guilt as a matter of having done something morally wrong (just like fear is about danger, shame about something disgraceful, and so on). The topic of collective emotions is a more general topic that does not only concern guilt but other emotions as well. A full account of guilt, just like a full account of any emotion, would have to include an explanation of how the emotion can take a collective form, but I am content to leave this to the theorists who specialise in collective emotion because this is an issue that requires a more

detailed study of not only guilt but how emotions, in general, can take a collective form. Also, since emotions in their individualistic forms seem developmentally prior (an infant does not start off from feeling fear for their mother but for themselves), I believe that it is also, in general, justified to start from emotions in their individualistic form, and later just adjust to make room for collective versions.

### 2.2.2. Self-disciplinary guilt

Next, let us consider what David Velleman (2003) has aptly labelled as self-disciplinary guilt. Cases of self-disciplinary guilt involve guilt over breaking a diet, not going to the gym or, more generally, not putting enough effort into a sporting activity, failing to do one's homework, and so on (Velleman, 2003; Maibom, 2019; Tangney et al., 1996). Allegedly, these are just private matters that fall short of true moral transgressions, so they are supposed to pose a problem to the idea that the norm violations that guilt concerns must necessarily be moral ones. This idea has been shared not only by psychologists (Tangney et al., 1996, p. 1257) but also by some philosophers (Harcourt, 2013, p. 139; Maibom, 2019, p. 42). Both seem to consider it as something obvious rather than something that needs to be defended. There are reasons for doubt though.

I want to start from the fact that not everyone feels guilt for not eating healthy and skipping exercise. One may either simply not care about those matters or may only feel regret rather than guilt. For example, one might regret having eaten too much because they feel really full now, and the food has made them weary. One may also regret their unhealthy eating habits in general once they notice that they have gained weight and realise that their favourite clothes do not fit them anymore. Neither of those cases needs to elicit any guilt feelings. This is because there is more to guilt feelings than simply thinking that, out of prudential reasons, you should have been more disciplined. So, what might it be that explains the shift from mere self-disciplinary regret to self-disciplinary guilt?

Let me begin with what I consider the most plausible explanation. I believe that, in most cases, the immorality of breaking a diet and skipping exercise is of an indirect kind. That is, one does not need to think that there is a general (perhaps Kantian or perfectionist) moral duty to stick to a particular diet and exercise.<sup>30</sup> Instead, the moral wrongfulness is mediated by promises: by breaking a diet or skipping exercise, one is breaking a promise to oneself and thereby letting oneself down. What is essential is that there be some previous, implicit or explicit, commitment to stick to a diet or exercise program. By means of promises, all kinds of nonmoral matters can become moral matters. There is no general moral duty to go to a movie theatre, but if you have promised to go with your friend, then you nevertheless have a moral duty to go to the movie theatre. And just like it is morally wrong to break a promise to another, it is morally wrong to break a promise to oneself.

---

<sup>30</sup> For an application of Kant's moral theory on that matter, see (Velleman, 2003).

Perhaps it could be suggested that breaking a promise to oneself is a less serious matter because the one to whom something was promised (the subject themselves), at least in some sense, consents to not following through with the promise. There is some intuitive appeal to this idea, but I think this appeal is only due to the fact that we are dealing with an unfair comparison. Rather than comparing breaking a promise to oneself with a paradigmatic case of breaking a promise to another where the other does not consent, we should compare it with those kinds of cases where the other does, at least in some sense, consent to breaking a promise. For example, imagine that you have a friend who has just gone through heartbreak and is afraid that their sorrow may make them overindulge with alcohol at the party where you were both planning to go. So, you promise your friend that you will keep your eye on them and will not let them drink too much. However, when partying, you both get carried away. You do mention to your friend that they were supposed to take it easy with the drinks, but as your friend is already under the influence, they say that let it be for just this once—seemingly consenting to you not following through with your promise. However, the next day, your friend feels really bad and wishes that you would have stopped them from drinking too much. I think that most would agree that even though your friend seemingly consented to you not following through with your promise, this does not fully undo the promise, and there is nevertheless a sense in which you have failed your moral duty. I think many self-disciplinary failures involve a similar dynamic: you make a promise to yourself to go to the gym, with the intention that this promise should be kept even if you do not feel like going to the gym, and then you do not feel like going to the gym, so you just let it go (seemingly consenting). However, later on, you regret. Even if cases with this kind of seeming consent were morally less serious than cases where the one to whom something was promised does not consent in any way, they are moral failures nonetheless.

Of course, none of this is to deny that there may be cases where the promisee can consent to the promise not being followed through, thereby freeing the promiser from the promise. Perhaps in some cases, the promise may have become irrelevant, or it might even be stupid to hang on to it, or the promisee might just change their mind about what they would like to oblige the promiser to do. But there is an important difference between those cases and cases of self-disciplinary failures, which were just further explained by the analogy of making a promise to a heartbroken friend that you will not let them drink too much. The crucial thing in the latter cases is that it is only in a very shallow and temporary sense that the person seems to withdraw from the promise. It just marks an episode of weakness of will (or drunken state of mind), not a permanent shift in one's priorities. And for this reason, one is not freed from the promise, and their guilt for breaking it will nevertheless involve a sense of moral wrongdoing.

The most common cases of guilt over self-disciplinary failures should thus be covered—cases where the duty to eat healthy and exercise is a localised duty that derives from a previous promise to oneself. The only alternative that is left for someone who wants to argue that guilt about eating unhealthy and shirking exercise is non-moral guilt is to argue that the subject has internalised a general

norm to eat healthy and work out, a norm that applies not only to them but other people as well, and simply disagree that such norm should be categorised as a moral one. This takes us to the more general territory of allegedly nonmoral norms, which goes beyond self-disciplinary norms that people often make promises about. For example, some of the allegedly nonmoral norms about the failure of which people are said to feel guilt include a prudential duty to visit the dentist, and duties of religious observance (Harcourt, 2013, p. 139). In this case, we are back to the debate between formal and material definitions of morality, and perhaps the somewhat blurry boundaries of morality. Presumably, the defendants of nonmoral guilt (Harcourt, 2013, p. 139; Maibom, 2019, p. 42) would want to adopt a material definition of morality, probably something along the classic lines that it must have to do with the well-being of other people. As already argued, I do not think that adopting a material definition of morality is a particularly promising strategy if our aim is to capture not just prototypical cases of guilt but also reach a more general definition that would cover more ground. It would rule out too many cases. If the traditional account of guilt is to be understood in its best form, then we should take it as referring to moral norms in a formal sense, and under this reading, it is no counter-argument to the theory that people feel guilt for violating a general norm to eat healthy or do exercise (or even more broadly, go to the dentist, and so on). It would probably just show that those kinds of norms are considered especially important in our current society, as they function in a way that is similar to how the more prototypical moral norms do (those that concern our behaviour towards other people). Alternatively, if we wanted to stick to a more material definition, we could just say that since there is no clear border between the moral and non-moral (an idea shared by many, for example (M. C. Nussbaum, 2001, p. 29)), we can regard cases of self-disciplinary guilt as borderline types of guilt, that is, as a kind of acceptable grey area that is only to be expected. For example, Patricia Greenspan has admitted that the rules that the subject violates when going off their health regimen are, in fact, nonmoral, yet she goes on to add that the subject nevertheless “treats them in emotional terms as if they had moral force” (1995, p. 164). So, on the one hand, in practice, they function in a similar manner. Yet, on the other hand, even though people do report guilt in such cases, this does not seem like a particularly serious form of guilt—something that people would, for example, confess on their deathbed. Compared to self-disciplinary guilt, guilt for merely causal harm and survivor guilt play a much more important and significant part in the grand narrative of people’s lives.

While both cases, collective and self-disciplinary guilt, may provide us with fruitful ground to discuss the smaller nuances of guilt and the somewhat fuzzy edges of the traditional conception, neither of them poses a serious threat to the unity of guilt in such a way as empathetic guilt and survivor guilt do. The latter two threaten to split the notion of guilt into different types that are relatively irreducible to each other. Emphasising empathetic guilt as a unique type of guilt that comes apart from traditional norm-violation guilt, Gangemi and Mancini (2011) have argued that it would be more appropriate to talk about guilts rather

than guilt. And Velleman (2003), when talking about the distinction between regular guilt and survivor guilt, suggests an account of guilt that is “disjunctive” rather than unified because one type of guilt has to do with loss of trust and another type with angry/envious resentment.

### 2.2.3. Guilt for merely causal harm

I will proceed with the two types of cases that will be the central focus of this dissertation and which are also the cases that Roy Baumeister and his colleagues have emphasised as key types of counter-examples to the traditional, intrapsychic, theory of guilt (1994, pp. 251–253). Those are cases of merely causal harm to others and guilt for unfair benefits (such as survivor guilt). Differently from collective guilt, those cases are more directly guilt-specific matters that concern the basic shape of guilt and, as such, pose an especially great problem to the idea that guilt is felt over one’s *morally wrongful deeds*. Differently from self-disciplinary guilt, those instances involve cases that are generally of much greater significance in the context of people’s lives as a whole. Furthermore, they do not just raise questions about how exactly we should draw a line between what counts as a moral norm and what counts as a nonmoral norm; they shed doubt on the very idea that guilt is about norm violations at all. Furthermore, it can even be argued that they threaten to split the notion of guilt into different types that are relatively irreducible to each other. Emphasising guilt for harm to others and survivor guilt as forming a unique type of guilt, altruistic guilt, that comes apart from transgression guilt and does not require norm violations, Gangemi and Mancini (2011) have argued that it would be more appropriate to talk about guilts rather than guilt. And Velleman (2003), when talking about the distinction between regular guilt and survivor guilt, suggests an account of guilt that is “disjunctive” rather than unified because one type of guilt has to do with loss of trust and another type with angry/envious resentment.

The cases of guilt for merely causal harm and survivor guilt will merit a longer discussion because I want to make it clear that there do not seem to be any judgments or thoughts of moral wrongdoing involved. I do want to note from the outset, though, that I do not expect those examples to convince the proponents of the traditional account that their account is wrong. As I will explain in section 2.3, ultimately the disagreement about those cases will lead to a meta-theoretical disagreement about which cases we want to include under our theory of guilt, and since our *explicandum* is different, there is an important sense in which we are simply engaged in separate projects that have a fundamentally different starting point.<sup>31</sup> But what this discussion of cases will help us achieve is that it will clarify which are the cases that are commonly spoken of under the label of guilt, but which the traditional account has serious difficulty accommodating. As such, it will help us map where exactly it is that common usage, which I want to do justice

---

<sup>31</sup> See more about the importance of a shared starting point, or the lack of it, from (Brun, 2022), especially section 4.2.

to, strays from the bounds of the traditional account. My suggestion for a potential proponent of the traditional account for reading those cases is just to be aware that what I talk about as guilt is something other than what the traditional account counts as guilt. I am not interested in those versions of those cases that do involve irrational beliefs or thoughts of wrongdoing, but those where there really seems to be no ground to attribute such beliefs. So, one should try to resist the urge to attribute irrational beliefs or thoughts, and just try to imagine how someone who does not think of themselves as morally blameworthy might nevertheless feel in those cases. All the potential responses to those cases will be addressed in the next bigger subsection (2.3), once the cases themselves have already been presented in their purified form.

Let us begin with cases of merely causal harm to others. In those cases, the subject has been part of the causal chain that led to harm to others but they have thereby done nothing morally wrong, and neither do they perceive themselves as such. Baumeister himself emphasises cases of unrequited love where one does not reciprocate the other person's romantic feelings (Baumeister et al., 1994). The subjects report guilt for having caused heartbreak to another but they do not seem to think that they have thereby done anything morally wrong. Indeed, it would be uncommon to think that there might be a duty to have romantic feelings for someone.

People who found themselves to be the objects of someone else's romantic attentions and did not reciprocate these feelings became the reluctant agents of the would-be lover's heartbreak. According to their accounts, they had done little or nothing to encourage the amorous attentions of the other person and were morally innocent, but they felt guilty afterward anyway. The relatively high rate of self-justifying statements in the rejectors' accounts likewise suggests that they struggled with guilt feelings that were not dispelled by their own arguments of technical, moral innocence. To be sure, people who had encouraged the would-be lover felt especially guilty, and so self-perceived responsibility may increase guilt; however, guilt feelings arose even in people who regarded themselves as not having done anything wrong. These findings suggest that the infliction of harm on another person, rather than any intentional or self-perceived wrongdoing, creates guilt. Moreover, they seem consistent with the hypothesized link between guilt and empathy, because people probably experience an empathic understanding of someone who loves them. (Baumeister et al., 1994, p. 253)

But there are also other cases: for example, such guilt is said to be experienced by careful drivers who have accidentally killed another person—for example, a child ran in their way, someone's bicycle swerved, or there was a motorcycle that appeared just behind the curve, driving on the wrong lane.<sup>32</sup>

When referring to cases of unrequited love and car accidents, I do not mean to deny that some of those cases may be explained in terms of the traditional

---

<sup>32</sup> The importance of those cases is also emphasised by Michael Zhao (2020). For interviews with drivers who have been involved in such accidents, see (Gregory, 2017).

theory that puts the focus on moral wrongdoing. There will likely be drivers who do think that they were morally at fault or are haunted by doubt that they may have been morally at fault. Similarly, those who have caused heartbreak are not all morally blameless—they may have encouraged or even exploited the other person’s romantic feelings (Baumeister & Wotman, 1994, p. 120, 136–137) and thus feel that they have acted in a morally wrong manner. Obviously, there is nothing that would prevent harm to others from resulting from morally wrongful actions, so it is only natural that some of those cases may involve a moral error, and that this moral error is also acknowledged by the subject. But this does not always seem to be the case. If a driver hit a motorcycle that appeared behind the curve and drove in the wrong lane, and everything happened so quickly that the person driving the car did not even understand what was going on (“I felt an impact, but a very strange impact. I thought maybe it was a deer” (Gregory, 2017)), then it seems highly incredible to say that they must think of themselves as being morally at fault. It is those kinds of cases of causal harm that I have in mind when I talk about *merely* causal harm.

In what follows, I will expand our understanding of merely causal harm by additional examples that help to make it especially evident that one need not see oneself as morally at fault. I apologise to the reader if some of those examples are of disturbing nature.

There are also cases of causal harm where there is no agency involved. In those cases, it is even more unlikely that the person might think that what happened was a result of their moral wrongdoing. For example, Michael Zhao invites us to think about the case of Elvis, who was born with a stillborn twin and later confided that he thinks he caused his twin to die while in the womb because he absorbed more than his share of nutrients (Zhao, 2020, p. 291).<sup>33</sup> Structurally, this case is also similar to how, during the Holocaust, wagons were so overloaded with prisoners that, inevitably, some ended up dying (Wiesel, 2006, p. 99; Lifton, 1991, p. 491; see more about those cases under 4.1). In both cases, the very fact that one took up valuable resources, be it either nutrients or mere physical space, meant that there was less for others, who ended up dying. Instances of causal harm that involve no agency can also include more sinister examples where someone is directly used as an instrument of atrocity. For example, psychologist Samuel Juni draws our attention to a scenario from Talmudic discussions, where “an immobilized victim is thrown on an infant resulting in the child’s death”. He comments: “I imagine that if I were forcefully bound up with rope and thrown on an infant, I would feel guilty about the murder—even though there was no way I could have prevented it. The mere fact that I was an instrument of the atrocity would incur guilt within my soul. Introspecting, it is clear that I would feel no

---

<sup>33</sup> Zhao presents this as a well-known case and does not provide a source. Some evidence can be found in Peter Guralnick’s biography of Elvis, though: “As a child Elvis was said to have frequently visited his brother’s grave; as an adult he referred to his twin again and again, reinforced by Gladys’ [her mother’s] belief that “when one twin dies, the one that lived got all the strength of both.” (Guralnick, 1994, p. 15)

culpability for the atrocity” (Juni, 2016, p. 3).<sup>34</sup> While such cases may be rare, there are also more daily examples where one’s body can make a difference to another living being’s life or death—cases of miscarriage that are in no way one’s own fault, and about which it has also been suggested that telling people that their guilt is irrational or misapplied does not really honour their sense that “something morally significant has been lost” (S. C. Miller, 2015, p. 152). All these examples go to show that people do not only care about what happens because of their moral agency or wilful choices, but they also care about the impact that their physical body as such can have on the world.

There are also cases where one is even further down the causal chain. Eger tells us of a Jewish man who went to a school entrance exam as a boy:

Sitting beside him at the exam was a boy named Adolf Hitler, who was not accepted at the school. But the Jewish boy was. And all his life, this man, who had left Europe and lived in Los Angeles, had felt guilty, because if Hitler hadn’t suffered this loss, if he hadn’t lost to a Jew, he might not have felt the need to scapegoat Jews. The Holocaust might not have happened. (Eger, 2017, p. 213)

Again, it seems highly unlikely, that throughout all his life, this man would have thought that he was morally at fault for passing the exam while Hitler failed. The focus is on the fact that he perceives himself as an important part of the causal chain that led up to the Holocaust.

Last but not least, it is also useful to note that contributing to a causal chain of events can take an even more indirect form. This has been well highlighted by Michael Zhao, who draws attention to a real-life story of a man who reported guilt for the fact that a firefighter had died in an attempt to rescue him (Zhao, 2020, p. 291). When driving over a bridge, the man had hit an icy patch and flipped over. Local firefighters were sent to rescue him from the wrecked car but, unfortunately, while trying to reach him, one of the firefighters slipped and fell through the gap between the highway’s east- and westbound lanes. (Garger et al., 2019) This is an example where someone wants to help the subject but we can also imagine indirect instances of guilt where someone seeks to bring about evil to the subject. Zhao invites us to imagine a case where an assassin accidentally shoots the subject’s loved one (Zhao, 2020, p. 294fn20). Of course, the harm need not be accidental. People can also feel causal guilt in cases where someone has sought to manipulate or avenge them by means of harming their loved ones. For example, had he had any emotional resources for that, Elie Wiesel could have felt guilt when his father was beaten by one of the kapos because Elie had not agreed to let the kapo have his golden crown (Wiesel, 2006, p. 55ff). Even though there is no doubt that the one to blame is the kapo, there is nevertheless a sense that it was

---

<sup>34</sup> Even though he talks about an imagined scenario, instances where someone was thrown on top of a pile of bodies and was able to feel others being crushed and suffocating below them, are an all too sad reality in many Holocaust stories (Wiesel, 2006, p. 93).

because of Elie that his father was beaten—he was the reason why the kapo turned his wrath on his father.

Overall, there are many ways how one can be a part of a causal chain that leads to harm to others. And people seem to consistently report guilt in those scenarios, even if there is little reason to think that they consider themselves as morally culpable.

#### 2.2.4. Survivor guilt

Let us now proceed with cases of survivor guilt. I want to start by noting that if survivor guilt is to be understood as any kind of guilt that a survivor might feel in relation to the traumatic events that they went through, then this is not the kind of survivor guilt that I have in mind here. As noted by psychotherapist Aphrodite Matsakis: “The death, injury, or insult of another, as is common in war or any type of traumatic experience, is a breeding ground for guilt” (1999, p. 20). In any catastrophe, there is usually an abundance of reasons why people can develop feelings of guilt. So, in what follows, I will try to disentangle some of the different senses in which survivors may feel guilt feelings, so that it would eventually become more clear what kind of guilt I have in mind. I will first try to explicate all those types of guilt that survivors may feel that can easily be accommodated by the traditional account. Even though those types of guilt are not my main concern in this dissertation, explicating them is necessary, so that we could already anticipate some of the ways how the proponents of the traditional account may try to write survivor guilt off as a rather traditional type of guilt. This discussion will eventually help us clarify what type of survivor guilt it is that poses an especially serious problem to the traditional account. This type will be the later focus of Chapter 4.

First, it is important to note that usually, many survivors of catastrophes feel a rather regular type of guilt for the egoistic actions that helped them stay alive. Tom Williams has labelled this kind of guilt content survivor guilt:

*Content guilt* [---] is a result of a person’s having done something to ensure his or her survival. This might include avoiding responding to others in need, having made a decision that resulted in others’ deaths, and seeking refuge for oneself when others remained threatened or suffering. This is a much easier form of survivor guilt to treat because the avoidant nature of this form of guilt implies a conscious decision to act, to survive, or operate efficaciously in the traumatic environment. (T. Williams, 1988, p. 323)

Even though some of this kind of guilt may be guilt for mere harm to others, much of content survivor guilt seems like the rather traditional kind of guilt that is to be explained in terms of perceived moral wrongdoing. Extreme catastrophe situations often make people more narrow-minded and egoistic because this is what helps them make it out alive. As put by Primo Levi (1988, p. 63): “the worst survived, the selfish, the violent, the insensitive, the collaborators of the ‘grey

zone', the spies. It was not a certain rule [---] but it was, nevertheless, a rule [---] The best all died." The same is emphasised in the context of the Hiroshima bombing as well:

In general, then, those who survived the atom bomb were the people who ignored their friends crying out *in extremis*; or who shook off wounded neighbors who clung to them, pleading to be saved [---] In short, those who survived the bomb were, if not merely lucky, in a greater or lesser degree selfish, self-centred, guided by instinct and not civilizations [---] and we know it, we who have survived. (Nagai 1951, 180–1; see also Lifton 1991, 48)

Both quotes draw attention to the fact that survival would have been nearly impossible if people had not been selfish. For this reason, many survivors feel guilt about their egoistic actions. This kind of guilt poses no counterexample to the traditional account of guilt, and thus, this is not the kind of guilt that I am concerned with.

Secondly, much of survivor guilt can be explained in terms of solidarity (sometimes also called separation)<sup>35</sup> guilt. This is the kind of guilt that one may feel when they break away from a group. For example, when someone moves out from their parents' house or breaks up with a relationship partner. It arises from the belief that one has violated the norm of sticking together with those they care about and sharing the same fate with them. How exactly does this play out in survivor guilt? In case one's loved ones did not make it out alive, one feels separated from them, and it feels as if they have violated the duty to stick together with them. Different authors have put this in slightly different terms. John Deigh (1999, p. 323) talks about it in terms of the requirement of friendship and comradeship to share in good times and bad. Fabrice Teroni and Otto Bruun (2011, p. 232), when referring back to Deigh, talk about it in terms of obligations of solidarity. Jordan MacKenzie and Michael Zhao (2023), too, prefer to talk in terms of solidarity. Patricia Greenspan (1995) talks about it as "a demand for "levelling": a requirement that the agent bring himself down to the level of others in a group with which he identifies" (p. 165). The idea behind all those formulations is that survivor guilt arises from a perceived duty to stick together with those one cares about and to share in both good and bad, which means that one is required to be at the same level with them—to share in their happiness, to share in their sadness, and not to exceed them in social competition or to otherwise

---

<sup>35</sup> This kind of guilt was first introduced under the label of 'separation' guilt by Arnold Modell, who reduced having a right to a life to having a right to a separate existence (1965). Weiss (1986) and Bush (1989) expanded this to include breaking away from a family by differing from them (see also O'Connor et al. 1997 p. 76). In both cases, breaking away is viewed in terms of loyalty and betrayal. Sometimes the labels of separation guilt and solidarity guilt are even combined. For example, O'Connor and colleagues talk about the guilt using the label "separation/disloyalty" guilt (1997). For my purposes, it is more clear to talk about solidarity because it gives a better clue with regard to what the normative ground is based on which the situation is viewed as morally bad.

come better off than they are (unless one shares those benefits that would otherwise separate him from his group). So, more specifically, there seem to be two duties embedded here: 1) to stick together with your group and not leave them, and 2) to share an equal position with them—both in emotional and material terms. It is important to note that at this point, the equality in question is not necessarily related to justice. There may be nothing unjust in working harder than others in your group and enjoying the benefits of your work. But one may nevertheless feel guilty about it insofar as they view themselves as having exceeded the norm of remaining equal with their peers. The key notion in play here is belonging to the group.

A sort of consensus seems to have emerged that solidarity guilt is indeed the main culprit when it comes to survivor guilt. At least two important researchers of survivor guilt have concluded that this is the core of survivor guilt—this includes Robert Jay Lifton (1991), who studied the survivors of Hiroshima, and eventually also William G. Niederland (1981), who was one of the first to introduce the term ‘survivor guilt’ to begin with (1961). But this form of guilt can easily be accounted for by the standard theory of guilt as involving transgression of moral norms (might be supererogatory norms as well, as long as they are important enough for the subject). Essentially, this kind of guilt takes the form of omission guilt—one feels guilty about not joining their loved ones in death, either at the time it would have been possible to, for example, accompany them to the gas chamber, or not sharing an equal position with them now in the present. In neither case does it pose a challenge to the standard conception.

Third, one may talk about survivor guilt as a kind of meta-emotion that one might feel for being glad to have survived while so many others died (Deigh, 1999, p. 323; Jäger & Bartsch, 2006). In this case, one perceives their joy as selfish and morally inappropriate. Partly, this may be mixed with solidarity guilt as well, as one may find it a breach of solidarity if they are happy about their survival and go about enjoying life, while their loved ones have died. But in any case, this kind of meta-emotion, too, can easily be construed as guilt of a rather traditional sort.

It is now time to move on to the kind of survivor guilt that I am concerned with. Tom Williams (1988) has labelled this kind of guilt as existential survivor guilt, which “is characterized by the survivor’s confusion over his or her having lived and the meaning of survival: “Why did I live when other people died?”” (p. 323) In this case, one feels guilt about the fact that they came out of a catastrophe alive, while others who were no worse than them did not. Usually, this takes place in the context of surviving a natural or human-caused catastrophe where lots of people are killed (e.g. floods, pandemics, wars, 9/11). But it can also be a response to a less wide-scale situation, like a car crash. In fact, survivor guilt can also go hand-in-hand with merely causal guilt. For example, in cases where one has caused a car accident that resulted in another person’s death, they may also feel survivor guilt on top of merely causal guilt. In his autobiographical self-help book David W. Peters notes: “I felt guilty for surviving when a young woman had been killed—by me” (Peters, 2023, p. 6). But it is important to note that the layer of survivor guilt is not due to one’s causal involvement in another

person's death but simply relates to the idea that one was no more deserving of surviving the accident than the other. As such, one can, of course, feel survivor guilt even if it was not them who was behind the wheel.

Many suggest that we should view this latter kind of survivor guilt as a form of a more general phenomenon that might be labelled "guilt over unjust enrichment" (Morris, 1987, p. 232) or "inequity guilt" (Baumeister et al., 1994, p. 248).<sup>36</sup> In addition to survivor guilt, guilt for unfair enrichment includes other cases as well. For example, there is a phenomenon called "Scandinavian guilt" that a person living in a welfare society may feel when they encounter a beggar. Or one of two siblings may feel guilt for being favoured over the other either by being born more healthy while the other has some chronic illness, or by receiving an unfairly big share of inheritance (Morris 1987: 233). The reason why those cases pose an especially great problem to the traditional account is that since there is no agency involved, it does not make sense to talk about any kind of wrongful action be it either morally so or in some other way. One simply feels guilt over a state of affairs that did not depend on his actions.

In their review, Baumeister and colleagues (1994) also list other cases where this type of guilt has arisen: among homosexual men who have tested negative for the human immunodeficiency virus, that is, who have been "spared at random", while others are tested positive and die (Wayment et al., 1995) and among people who keep their jobs while their colleagues are fired (Brockner et al., 1985).

Guilt for unfair enrichment is often confused with solidarity guilt. For example, Nancy Sherman seems to talk about a type of guilt which is very similar to guilt over unjust enrichment,<sup>37</sup> which she introduces as a "generalized form of "survivor guilt"" under the label of "luck guilt" (Sherman, 2013, p. 185). Luck guilt is supposed to apply to a situation where it is simply "a matter of dumb luck that a soldier makes it out alive, and tragic luck that others do not" (pp. 185–186). Since it is just a matter of luck, it is clear that fate is not distributed based on merit and fairness, which is the reason why I mention this here. Though it should be noted that Sherman herself does not really explicitly talk about it with reference to justice. Yet the guilt that she talks about is not really guilt about luck as such (or, it is a matter of luck only from the outside perspective (p. 190)) because ultimately she reduces it into a felt "betrayal of solidarity". She considers this kind of guilt virtuous but "strictly speaking irrational" because, from the position of a fair judge (the outside view), the person has done nothing wrong (p. 192). Also, John Deigh reduced inequity guilt to the requirement of friendship and comradeship to share both in good times and bad (Deigh, 1999, p. 323).

But solidarity guilt is not the same as guilt for unjust privileges as such. To be precise, the guilt that Deigh and Sherman talk about does not seem to have so much to do with justice but with the duty of solidarity, which is more closely

---

<sup>36</sup> Even though, as we will see in 4.3.1, he may have some reservations about that label, as he does not think that inequity alone is sufficient (Baumeister et al., 1994, p. 248).

<sup>37</sup> Sherman's second example doesn't really seem like inequity guilt.

related to belonging to a group. As such, it is actually better described as solidarity guilt. In fact, authors often seem to confuse the two, and I have never seen those two types of guilt properly disentangled. But it is important to differentiate between the two because their intentional focus is different: one focuses on group belonging, and the other focuses on justice. In great part, cases of inequity guilt and solidarity guilt co-occur, as cases of guilt over unjust enrichment often involve people with whom one belongs to the same group and feels a special sense of closeness. But inequity guilt may also arise in comparison with strangers. Of course, one may feel solidarity with strangers as well, but what is important here is that the root of guilt lies in the injustice of the situation—in the fact that one was no worthier or even less worthy of the benefit than the other(s). While injustices may threaten one's group belonging and thus often exist in combination with separation guilt, one need not be concerned about justice *per se* because they are worried that this might pose a threat to their group belonging. Concern for justice may also exist as an independent concern, and this is also the reason why it becomes more evident when looking at cases dealing with strangers, with whom the identificatory ties are somewhat looser.

Having set aside cases that are better described as solidarity guilt, I would like to point out a further way how one might try to reduce cases arising from true inequity to regular guilt (though I will not be arguing that this move is necessarily justified). While it is not the survivor's fault that they came out as winners in the unfair roulette of life, the situation of injustice can be taken to give rise to the duty to rectify injustice. As such, it might be suggested that the guilt that survivors feel is not the classic kind of action guilt that requires that one must have done something wrong. Instead, it is a form of omission guilt (which may well render its object slightly more invisible and explain why it is that it may seem that people feel guilt about the mere fact of being alive, even though actually it is about the failure to balance the injustices). On the background of more everyday situations, this seems to make sense (one might well feel guilty if they received more than their fair share, and the guilt would result from them taking/accepting it and not sharing it, which are all actions done by the subject). Consider little Amanda who gets a bag of candies all to herself. Insofar as she does not consider herself more worthy to receive it than other children, she may feel a duty to share it with others and thus be susceptible to guilt unless she does so. Survivors may implicitly feel a similar kind of duty—a duty to share their good fortune with others. The tragedy of the survival situation lies in the fact that there is nothing to be done about the matter, life cannot be redistributed.

This kind of reduction of inequity guilt to omission guilt has been proposed by various authors (Prinz & Nichols, 2010). In fact, the possibility that survivor guilt might be reducible to guilt over the violation to rectify fortuitous imbalances or injustices has also been pointed out by Herbert Morris, though he thinks that the true essence of survivor guilt goes deeper than that and lies in the mere state of having unfairly benefitted. He invites us to imagine a survivor who might describe their feelings as follows:

It has been difficult for me to accept that I preferred my friend's death to my own when one of us had to die. I even think that my feeling guilty is partly explained by that preference and certain odd, unconscious, beliefs I had about my responsibility for his death. I also realize that I feel some guilt over not doing all that I reasonably could to show my gratitude for my good fortune. But when all those bases are covered, I still seem to feel guilty just over the fact that I profited from something I see as unfair, that I should have been miraculously saved and not my friend. (Morris, 1987, p. 236)

While acknowledging that there are indeed many grounds for guilt in the survival situation, Morris suggests that if we subtract all other grounds, there will still be some guilt left for the mere fact that one survived while the others, who were equally worthy, did not. So, the object of guilt would be the state of being unfairly benefitted. Talking about guilt in the objective state, he points out that if this was the case, then “[o]ne's obligations would follow from one's guilt, not one's guilt from unfulfilled obligations” (Morris, 1987, p. 236). By those words, he wants to emphasise that it is not because of unfulfilled obligations that one can be considered guilty, but rather because of the state of guilt that one acquires the obligations to rectify the imbalance. In other words, guilt comes before the duty to rectify injustices.

It is this type of survivor guilt that poses the greatest problem for the standard account of guilt, and which is philosophically the most interesting one. It is also this type of guilt for unjust enrichment that is linked to the “Why me?” question that survivors are often haunted by. Even though some have tried to reduce guilt over unjust enrichment to solidarity guilt, solidarity guilt would not really explain the “Why me?” question that survivors are often haunted by. As pointed out by MacKenzie and Zhao (2023, p. 2721), someone who feels solidarity guilt may think of their privileged position as well deserved, yet still feel guilt because their better status in life estranges them from loved ones. In case of true guilt over unjust enrichment, the guilt is not felt over additional complications that this unfair enrichment may cause but about the sheer fact that the enrichment is unfair.

It is never impossible that the subject may have irrational beliefs about having done something morally wrong, and one may always have concerns about isolated counter-examples. But I hope that, in conjunction, the cases discussed above nevertheless pose a strong cumulative case against the traditional account of guilt. Later on in the dissertation, I will take this case only to be strengthened because, eventually, I will show that all those cases are systematically pointing towards a different kind of way how one might construe the shape of guilt—a way that is unified with reference to moral badness rather than moral wrongness.

### **2.3. The traditionalist response to the problem cases**

Given that we have already presented the counter cases in their strongest form, as cases where there does not seem to be any kind of judgment or thoughts of moral wrongdoing involved, there are basically only two ways how the

proponents of the traditional view can respond. First, they would either have to keep on insisting that, at some deeper, perhaps unconscious level, these cases must nevertheless involve judgments or thoughts of wrongdoing. Or, secondly, they can accept that there is no perceived moral wrongdoing and argue that those cases are better described as something other than guilt—perhaps remorse, regret, agent-regret, or something alike. In this section, I will review and offer a critical response to both strategies.

### 2.3.1. Attribution of irrational judgments or thoughts

I have previously listed many cases where it seems highly unlikely that the person must necessarily believe or think that they have done something morally wrongful. Especially in cases where there is no agency involved at all—for example, the guilt felt about the firefighter losing their life in an attempt to save the subject, the man who felt guilt about outdoing Hitler at an entrance exam, cases where one's body is used as an instrument of evil, existential survivor guilt where one is simply burdened by the question “Why me?”, for example, the kind of guilt that a sole survivor of a plane crash might feel. It has not been my wish to deny that there may, of course, be cases where people do hold irrational beliefs and think of themselves as morally at fault, but this does not seem to be necessarily the case and in some cases, the subjects even explicitly say that they do not think that they are morally at fault (Baumeister & Wotman, 1994; Gregory, 2017).

To be fair, the fact that people report the very opposite beliefs has not necessarily been considered a problem for the traditionalists. It could be suggested that even if one has the belief that they are morally innocent, they may also have other, contradictory beliefs, according to which they are nevertheless guilty of moral wrongdoing. As pointed out by John Deigh (2019), we might view those people as having a divided mind: ““Part of me,” the person might say, “believes that I’m guilty even though I don’t believe it” (p. 69 fn 14). Such cases could be explained by Donald Davidson’s (2004) idea, according to which we have to distinguish between accepting a contradictory proposition and accepting separately two contrary propositions that are held apart and which continue to exist in our mind embedded in semi-independent structures of interlocking beliefs, desires, emotions, memories, and other mental states. Irrationality tends to get a bad rep, but this is only because we tend to think of it as something unusual. So, if someone is deemed irrational, it seems to imply that, compared to other, normal, people, there is something wrong with them. But this need not be so. People tend to be irrational fairly often. We are complicated creatures, and it would be overly naïve to think that our mental life is characterised by full logical consistency. Once we accept this, there is no special charge in characterising some of our beliefs as contradictory or irrational.

But it would nevertheless be extremely uncharitable to insist that this must always be the case. If, from a certain point onward, the proponent of the traditional view still kept on insisting that one must nevertheless have unconscious

beliefs<sup>38</sup> about being morally at fault, their very stance turns into an irrefutable dogma. The theory would cut itself off from the best kind of empirical confirmation that it might have—self-reports about what people actually believe. Once self-reports are not given due credibility anymore, it will become unclear how we could empirically verify whether the person takes themselves as morally at fault. Notice that since, according to the cognitivist approach, emotions are to be explained in terms of beliefs, one cannot just take the opposite strategy and attribute beliefs on the basis that one simply has the relevant emotion. That would be circular.

The problem of becoming an irrefutable dogma is a serious one. This would basically turn the traditional approach into a pseudoscientific theory that cannot be refuted. As such, it would be similar to Freud's idea that behind every psychological problem is the Oedipal complex or that every dream is a wish fulfilment. Whenever someone would try to prove the opposite, one could just say that they have repressed their Oedipal complex or that the dream that does not seem like wish fulfilment actually just fulfils one's wish to prove that every dream is not wish fulfilment. In modern science, we cannot operate with these kinds of irrefutable dogmas. As such, the proponent of the traditional approach would do well not to insist, at least from a certain point onward, that there must be unconscious irrational beliefs involved. And the kind of purified counter-examples represent precisely this point—given that there is neither any ground for such beliefs as it is often not even one's agency that is involved, and also no evidence in support of the idea that the subject has irrational beliefs, it would simply be both uncharitable and overly dogmatic to continue with such belief.

### 2.3.2. Denying the status of guilt

In cases where the subject has no thoughts or judgments of being morally at fault, the only option left for the traditional view is to say that this is not an instance of true guilt and should instead be described as some other emotion.

This is the strategy employed by one of the leading proponents of the traditional view, John Deigh. Even though he admits that how we choose to label our emotion concepts is ultimately "a matter of choice" (Deigh, 1982, p. 399), his main rationale for reserving 'guilt' for only those cases where one does believe that they have violated a moral norm, is preserving the tight link between the subjective feeling of guilt and the objective state of guilt: the feeling of guilt implies that one also perceives themselves as objectively guilty, i.e. responsible for moral wrongdoing. On the one hand, this is common sense. We use the same word for both the objective state of being guilty and the subjective feeling. This does not seem like a coincidence. While the word *lie*, for instance, is a homonym,

---

<sup>38</sup> I take it that the notion of unconscious thoughts would be a conceptual impossibility because, differently from beliefs, which are dispositional (you can be said to believe that the world is round even if you do not think about it all the time), thoughts are by their nature episodic—they are parts of our stream of consciousness.

which has two totally different meanings—one that refers to reclining, and the other that refers to telling a falsehood –, the two meanings of guilt refer to phenomena that are tightly connected and often seem to go hand in hand. Therefore, taking the belief that one is guilty as a conceptual fixed point to anchor our understanding of the emotion of guilt seems like a *prima facie* reasonable thing to do. This is not to say that we should always naïvely rely on language, for language can also be deceptive. But it does give us an initial reason to look for commonalities rather than suspect that the similarity between the words is just a coincidence. In case of being insulted and feeling insulted we do not suspect that feeling insulted might be about something altogether different than about believing that one is insulted. So, why should we think so when it comes to being guilty and feeling guilty? All-in-all, there seems to be a strong *prima facie* linguistic argument in favour of keeping the subjective feeling of guilt linked to the (perceived) objective state of guilt. On the other hand, given that Deigh works at the intersection of philosophy and law, his reasons for considering the state of guilt as a conceptual fixed point may be even stronger: guilt has juridical implications and is associated with punishment (Deigh, 1982, p. 399). People who feel guilt without perceived wrongdoing are often just victims of unfortunate circumstances and introducing a word that has associations with objective guilt and punishment is inappropriate.

So, extending the word ‘guilt’ to cases without perceived wrongdoing raises concerns about victim blaming. To some extent, those concerns are probably somewhat inflated. Even the traditional view that does tie guilt to perceived wrongdoing does not necessarily imply that if one feels guilt then others have the right to blame them. Patricia Greenspan (1995) emphasises that a person may have a right to hold oneself accountable to higher demands on morality than others. Fabrice Teroni and Otto Bruun (2011) add that even if it is irrational to hold oneself responsible to such a higher standard, it may nevertheless be a sign of virtuous character. It is, of course, true that the traditional view nevertheless implies that the subject herself has the right to blame themselves or that blaming oneself may be a sign of virtuous character. But when we talk about the more liberal practice of extending the word ‘guilt’ to cases without any kind of perceived wrongdoing, then there really is no such implication. The proponents of the more liberal view are quite explicit in suggesting that we break the connection between the feeling of guilt and the objective state of guilt: feeling guilt is not a matter of perceiving oneself as bearing the objective state of guilt.

This is an important point, so let me iterate it in slightly different terms as well. To the extent that being in an objective state of guilt is synonymous with being *guilty*, this also means that the proponents of a more liberal view, myself included, are not talking in terms of the subject *feeling guilty*—a phrase that seems to come naturally in light of the traditional view that often functions by default, but does not really strike a chord with the more liberal view that detaches guilt from wrongdoing. For some people, at least, such associations are hard to root out. In the hopes that this will help the reader to better remember this point,

let it be mentioned that throughout many presentations about guilt without wrongdoing, I still sometimes receive sceptical questions that go along the lines “So, you are saying that feeling guilty is about ...?” It is no wonder that a more liberal account raises scepticism and makes little sense if it is interpreted as an account of feeling *guilty*. The very assumption that I am talking about feeling *guilty* means that I have not been heard. The concept of guilt that I am interested in is detached from notions like wrongdoing, the objective state of guilt, or being guilty. Understanding guilt without wrongdoing may require a paradigm shift, and it might take conscious effort to set aside associations with the traditional account of guilt.

But even if no one actually means to suggest that the practice of talking about guilt in a wider sense should have any implications of victim blaming, talking in terms of guilt may be harmful precisely because it introduces the wrong kind of associations with objective guilt.<sup>39</sup> As pointed out by Wendy S. Harpham, a doctor of internal medicine, cancer survivor and author of many books about cancer and healing: “It’s sad and unsettling for survivors when other patients die. Calling the feeling “survivor guilt” unnecessarily adds to their discomfort. For now, unless you have a better suggestion, let’s try calling it “sadtude” to help survivors adjust to the news and choose an attitude of gratitude and hope.”” (Harpham, 2018) In more philosophical terms, what Harpham suggests is that we opt for a more radical brand of conceptual engineering and label our concepts in a way that would have beneficial social effects. It is interesting to note that if we renamed all survivor guilt as ‘sadtude’, an amalgam of sadness and gratitude, then this label would not necessarily capture the original phenomenology of emotion. Even though it gets the sadness aspect right, gratitude may not be the dominating element of the emotion—people who report survivor guilt are often much more focused on the injustice of being spared. But the idea is that, hopefully, talking in terms of ‘sadtude’ would steer their emotions in a more positive direction.

As it can be seen, there are many considerations regarding how we should label our emotion concepts, and taking a stand on what is an all-things-considered best label is a complicated, and ultimately also a normative, not just a descriptive, matter. This dissertation is not meant as a final say on how we should use the word ‘guilt’ or which label is the best to capture cases without moral wrongdoing. I am more interested in the descriptive question that concerns what kind of rationale lies behind the common practice of extending the label of guilt to cases without moral wrongdoing. In that sense, I do not mind if proponents of the traditional account, like Deigh, would reserve ‘guilt’ for only those cases that involve perceived wrongdoing, and call the wider emotion genus by some other name. Because ultimately, this dissertation is not about labels. The aim is to gain a better understanding of what is going on in those alleged cases of guilt without wrongdoing, and in what kind of relationship they stand to the traditional instances of guilt.

Yet if I am going to speak about those cases, I will inevitably have to call them by some name. And regardless of how I choose to call those cases (guilt or not

---

<sup>39</sup> This danger has also been acknowledged in (Lifton, 1996, p. 144).

guilt), I will sound more sympathetic to either the moral traditional conception of guilt or the more liberal and inclusive practice. So, if I have to pick a side, then I would prefer to be on the inclusive side. Partly, this is just a practical matter. Since both the folk practice that I am interested in and also all the theories that I am concerned with in the rest of this dissertation opt for a more inclusive understanding of guilt, it would simply be inconvenient and redundantly revisionary to always translate their claims into a different terminology. But there are also some more substantive reasons behind extending guilt to a broader range of cases. Even though I do not expect them to settle the question of what is the all-things-considered best label for the wider emotion genus that I am concerned with, they should nevertheless be viewed as strong reasons in favour of calling it guilt. The most important consideration lies in the simple fact that there is already a relatively widespread practice of labelling the phenomenon as guilt. While the traditional account takes the *prima facie* linguistic link with the objective state of guilt as a conceptual fixed point, my own more liberal use of the word is anchored in a relatively common folk use. It is not clear that holding on to some linguistic link is necessarily more important than being aligned with how many people actually talk about guilt. On the one hand, I think it is important to respect how people talk about emotions. If we are not going to take self-report seriously, then it is hard to see how we can go about researching more complicated emotions at all—emotions like guilt, pride, jealousy, as opposed to simpler emotions like fear that can be more easily explained in terms of physiological changes. If we want to continue to talk about emotions in first personal terms (as has been so common in philosophy), then suggesting that we ignore people’s self-reports would be self-defeating because it would cut us off from one of the most important sources of evidence for our theories. This does not mean that we have to accommodate how each and every person talks about guilt. Admittedly, some people may use emotion terms in idiosyncratic ways. But, as also emphasised by Herbert Morris, there is reason to be sceptical about “any claim of widespread misuse of terms for emotional states” (Morris 1987, 221). On the other hand, I am also somewhat pessimistic about philosophical attempts to overcorrect ordinary language. John Deigh might have suggested already back in 1982 that we resurrect the term ‘remorse’ and put it into good use when talking about the emotional experience that is felt in response to causing harm to others (Deigh, 1982, pp. 399–400), but little seems to have changed since then. If anything at all, then it seems that it has become even more common to talk about guilt in wider terms—we have prominent psychological theories that view guilt in terms of harm to others, and terms like ‘survivor guilt’ are a common part of ordinary language. So, even if some may consider it a bad thing, it seems that the wider use of guilt is here to stay. Last but not least, I would like to emphasise that it is not clear to many that any alternative would be considerably better. The distinction that some philosophers make between ‘guilt’ and ‘remorse’ is explicitly proposed as a technical division adopted for philosophical purposes (Deigh, 1982; Taylor, 1985). It does not reflect how people commonly use these terms: when talking about cases without wrongdoing people nevertheless tend to talk about guilt rather than remorse. Not

as a decisive reason but as a remark, I would also like to add that in my native language, Estonian, there are no separate words for the two, and remorse is very hard to translate into something other than guilt. In addition to remorse, some might suggest that, at least in cases of merely causal harm, it might be appropriate to talk about Bernard Williams' (1976) notion of agent-regret. Again, this is a technical term that is even less popular in everyday language than remorse. I agree that agent-regret is relevant for cases of merely causal harm, yet my problem with the term is that, with regard to the phenomena that I want to explain, it stops halfway.<sup>40</sup> In some cases of causing harm, it may indeed be our agency that is involved (e.g. us wilfully driving on a certain road). Yet there are also cases without any agency. Remember the examples where one's body is simply used as an instrument of evil and thrown on top of others, who will suffer as a result. People do not only care about the impact that their agency has on the world, they also care about the sheer impact that their physical body may have. And, besides, even though interpreting Williams is an exegetical matter for which there is no room in this dissertation, I would like to highlight that Williams himself was not opposed to saying that agent-regret is a form of guilt. He admits that agent-regret "can be psychologically and structurally a manifestation of guilt" (B. Williams, 1993, pp. 92–93). But if agent-regret is too narrow, then perhaps we could just talk about regret in general? My problem with the latter suggestion is that regret is too broad of a term that does not really capture the moral associations that, as I will be showing by the end of this dissertation, underlie the phenomenon. People feel regret about all kinds of things, many of which are of totally non-moral kind—buying a polyester dress that is not particularly breathable, not taking a water bottle with you, etc. For this reason, I find it perfectly natural that the folk practice gravitates towards grouping cases of merely causal harm and survivor guilt alongside guilt rather than regret. Saying that you regret having killed another person in a car accident just does not sound right because it does not capture the moral seriousness of the matter. As I will argue, even when guilt

---

<sup>40</sup> This remark applies to agent-regret understood as a kind of regret that one feels for the bad results of their actions (Bagnoli, 2000; Gardner, 2018; Wolf, 2001). Interestingly, Jake Wojtowicz has noted that agent-regret has not really received a particularly thorough treatment and remains a misunderstood emotion (Wojtowicz, 2019, p. 9). He (2019, 2022) argues that agent-regret is not first and foremost a matter of bad result but is instead directed at the fact that one's own agency was involved in bringing about the result (which the subject might otherwise salute). This is supported by the consideration of "pure cases" where one just regrets one's action but does not regret the result. For example, a rifleman in a firing squad may wish to see justice served and the convict executed, yet still regret that it was he who killed the convict. If his analysis is correct, then this might be all the better for agent-regret, though. As I have just pointed out, agent-regret, at least how it has been usually understood, seems to stop halfway—at those cases where harm is caused by agency, whereas guilt for harm to others can be brought about non-agential failures as well—it might sound somewhat arbitrary to make agent-regret a separate emotion type rather than just a subtype of guilt. However, if Wojtowicz is right and the focus is not on the bad end result (and the subject need not even view the end result as bad) but the emotion is focused on one's agency to begin with, then it stands a stronger chance for contending to be a truly unique emotion on its own.

is not felt about moral wrongdoing, it is nevertheless linked with morally bad states of affairs, so I hope that the connections with morality will become clearer as this dissertation proceeds.

## 2.4. How to move forward?

If the traditional account is unsatisfying, then how exactly should we go about improving our account of guilt? Before making my own suggestion, I would first like to take a look at an alternative strategy. Namely, it might be suggested that the problem of the standard account lies in its cognitivism: once we give up the requirement of belief (regardless of its content), the problem disappears. The proponents of the feeling theory would suggest that we identify guilt with a specific phenomenal feel, and the proponents of the motivational theory would put an emphasis on the action tendency of apologising and making amends. In what follows, I will argue that, in the debate that I am concerned with, focusing on the theory of emotions is not the most important matter. The clash between the standard account of guilt and the cases without wrongdoing can, at a more fundamental level, be reformulated so that it is not necessarily the *belief or thought* of moral wrongdoing that does not fit well with cases without perceived wrongdoing but the very idea that the *core relational theme* of guilt must be that of moral wrongdoing. This will provide an emotion theory neutral framework for the rest of the dissertation that should be sufficiently attractive for a relatively wide audience. Even though the idea that we can have meaningful discussions about the nature of specific emotion types without specifying the emotion theory is rarely elaborated on, this is also the approach that Jesse Prinz and Shaun Nichols (2010, pp. 119, 133) explicitly adopt for their discussion of guilt, and I suspect that this is also the approach that tacitly underlies many other discussions as well. For example, when Baumeister proposes that instead of viewing guilt as a matter of moral norm violations but harm to others (Baumeister et al., 1994), he never explicitly elaborates on the intricacies of different emotion theories, but seems much more concerned with the more simple but also more guilt-specific question of what guilt is ultimately about—what is the core relational theme of guilt.

I will begin by first laying a theoretical foundation by introducing the notion of core relational theme and clarifying how, according to different theories of emotion, each emotion type represents<sup>41</sup> a distinct core relational theme. Then, I will move on to show why changing the emotion theory alone, while leaving the core relational theme of moral wrongdoing intact, cannot be sufficient to do justice to cases without perceived wrongdoing.

---

<sup>41</sup> I have relied on a widely shared assumption here that emotions are intentional and that their intentionality is to be explained in terms of their representational content, which I think is a fairly unproblematic assumption, especially when it comes to such a complex emotion as guilt. However, for those interested in how the notion of core relational theme can be integrated into a “truly enactivist” non-representationalist framework, see Hutto (2012).

The term *core relational theme* was first introduced by Richard Lazarus (1991) to denote the “ongoing relationships between persons and their environments” that emotions represent (p. 22). It is the relationship between the individual and the world that is of importance here—emotions mark the significance that certain states of affairs have to the subject’s well-being. For example, it might be said that anger represents offence, fear—danger, sadness—loss.<sup>42</sup> According to the traditional account of guilt, the core relational theme of guilt is one’s own moral wrongdoing. Some may also prefer to talk in terms of “core evaluative concern (or theme)” (Ben-Ze’ev, 2002, p. 148), but I will stick to the core relational theme, as this has become a more standard phrase.

But how exactly is the core relational theme related to the structure of emotion? The answer is going to depend on one’s theory of emotion.<sup>43</sup> For the sake of simplicity, I am going to illustrate how the notion can be tied with the three main theories of emotions in philosophy—cognitivism, feeling theory, and motivational theory.

Cognitivism offers the most simple and straightforward account of how emotions represent their core relational themes—they involve an evaluative cognition (judgment or thought) about the particular object of the emotion instantiating the core relational theme of the relevant emotion type. For example, when you are afraid, you judge the dog that is chasing you to instantiate a danger. Or, when you feel guilt, you judge your deed as morally wrong (at least according to the traditional view). In this framework, core relational theme is reflected in the inner structure of the emotion—it is mirrored in the very content of the propositional attitude that constitutes the emotion.

Some authors who write about guilt, for example, Michael Zhao, seem to use the words ‘cognitive’ and ‘representational’ synonymously (Zhao, 2020, p. 286). This ignores the possibility that something can be representational without at the same time being cognitive. However, the way how other theories of emotion can talk about emotions as representing the instantiation of the core relational theme is slightly more round-about and involves making use of the teleosemantic account of representation. I will first explain this strategy on the example of the feeling theory. According to one of the most sophisticated versions of the feeling theory by Jesse Prinz (2004), emotions are constituted by certain felt bodily

---

<sup>42</sup> In an alternative terminology, offense, danger, and loss have also been described as the *formal objects* of the corresponding emotion types (Kenny, 1963), or the main determinants of the *shape* of the emotion (D’Arms & Jacobson, 2000).

<sup>43</sup> Lazarus himself defended a “cognitive-relational-motivational” (1991, p. 125) theory of emotion that has ties with both cognitive and motivational theories. Even though Lazarus is usually considered an appraisal theorist, and appraisal theories are often taken as counterparts of the philosophical cognitivist theories within the field of psychology, it has also been suggested that his theory fits best into the motivational tradition (Scarantino & de Sousa, 2018, p. 19). I take it that his ties with different theories of emotion partly help to explain why the notion of core relational theme has received such a wide resonance. But this is not the place for an exegesis of Lazarus’s own account. What is important for the current purposes is how the notion of core relational theme functions in a wider discourse.

changes that represent core relational themes in light of Fred Dretske's (1986) teleosemantic account of representation. Under this interpretation, to say that the bodily changes (that constitute the emotion) represent a core relational theme means that those bodily changes are reliably caused and set up (by learning or evolution) to detect a particular core relational theme. Prinz illustrates this with the analogy of how a simple beep emitted by a fuzz buster can represent the presence of a police radar because it is set up for that purpose. This helps us better understand how something quite simple (the beep) can represent something quite complex (the presence of a police radar). It helps us understand that in order for the beep to represent the presence of a police radar, the presence of police radars does not have to be mirrored in the inner structure of the beep. Similarly, the core relational theme does not have to be mirrored in the inner structure of the bodily changes that constitute the emotion, yet the bodily changes can nevertheless represent the same thing as the corresponding belief. (Prinz, 2004, pp. 65–66) As such, emotions have an important epistemic function: they “use our bodies to tell us how we are faring in the world” (p. 69). Under this interpretation, the judgments that go along with different emotion types are not really the causes of emotion (as in the case of the cognitive theory) but rather something that may be formed precisely on the basis of the emotion. The emotions are like little beeps of the fuzz buster that let us know about important changes in our relationship with our environment. But the most important thing is that even if one does not form a judgment on the basis of emotion, their emotion nevertheless represents the situation as exemplifying a certain core relational theme.

Let us now move on to the motivational theory of emotion. The proponents of the motivational theory put the focus on the motivational aspect of emotions like states of action readiness, action tendencies, or goals. For the sake of simplicity, I will continue to talk in terms of action tendency alone. To the extent that the focus is on the action tendency, the motivational theory of emotion can allow different tokens of an emotion type have a different phenomenal feel as long as they make us act in a similar way. But the way that the motivational approach can account for the idea that emotions represent core relational themes is basically the same teleosemantic approach adopted by Prinz, according to which emotions represent core relational themes in virtue of having the function of being elicited by them (Scarantino, 2014, p. 178). Only here, since the constitutive element is the action tendency rather than the phenomenal feel, it is the latter that is the bearer of this function: the action tendency is set up to be elicited by the presence of certain core relational themes. According to Andrea Scarantino, this reflects the informational (epistemic) function of emotions. He would only add that in addition to the informational function, emotions also have a motivational function—the action tendency is not only meant to signal the presence of a core relational theme but also serves the function of making us respond to the core relational theme in a way that helps us achieve certain relational goals.

This concludes the most dominant approaches in emotion theory. What is common to them all is that they consider emotions as linked to their core relational themes via the fact that emotions, in one way or another, represent the

instantiation of the core relational theme. I take it that most theorists should already be satisfied by one of the options outlined above, and I will myself continue to talk in terms of representation. So, when I say that the subject *perceives* oneself as violating a moral norm, all I mean is that they take the emotion to represent that they have violated a moral norm.

However, there is an important caveat that I would like to elaborate on. Saying that emotions represent core relational themes, which are usually specified in terms of evaluative properties like danger, shameful, moral wrongness, etc., is usually taken to offer us a standard based on which we can assess whether guilt is fitting or not. Justin D’Arms (2022) has called such views alethic because they suggest that the fittingness of the emotion is a matter of representational accuracy—once we know what the core relational theme of the emotion type is about, we can just check whether the object of the particular emotion token instantiates the core relational theme or not. However, the idea that analyses of core relational themes may provide us with such an independent standard may be an illusion. If we follow the sentimentalist tradition, which views those evaluative properties as themselves response-dependent (i.e., to be explained in terms of the relevant emotions), then there is a sense in which saying that someone perceives a dog as dangerous amounts to little over and above the fact that they are afraid of the dog. If the sentimentalist fitting attitude analysis of value is correct, then the whole talk about emotion representing certain core relational themes seems circular and not particularly informative.

The alternative that D’Arms (2022) offers to the alethic views is what he calls ‘interpretivism’.<sup>44</sup> Even though he has his own reasons for combining the view with the motivational theory of emotion, it could, in principle, be combined with the feeling theory as well. But it is incompatible with cognitivism: since emotion is explanatorily prior to the evaluative thought, we cannot explain the emotion in terms of the thought. According to the interpretivist view, the talk about core relational themes would not be explanatorily prior to the talk about emotions. The talk about the core relational theme simply offers an interpretive “gloss of how someone in the grip of a natural emotion is best understood as appraising its object” (D’Arms & Jacobson, 2023, p. 142), it is “an interpretation of the nature of an emotion as a whole” (D’Arms, 2022, p. 118). I think that the metaphor of translating would be useful here: we are basically trying to translate the gestalt of the emotional experience (be it either more feeling-based or motivation-based) into words, and we are trying to align it with our verbal discourse about values. As with any translation, there is a sense in which it will be imprecise. This is also acknowledged by D’Arms and Jacobson (2023): “Since emotional appraisals are response-dependent, any gloss of them in ordinary language will be a rough-and-ready interpretation, and inevitably imprecise insofar as it is given in language

---

<sup>44</sup> This kind of ‘interpretivism’ should not be confused with how interpretivism is understood in the philosophy of mind—as a view, on which what mental states one has crucially depends on interpretation rather than one-to-one correspondence with some underlying interpretation-independent reality (Dennett, 1987; Byrne, 1998; Mölder, 2010).

that does not refer to the response itself” (p. 137). But what is important for the current purposes is that even sentimentalists like D’Arms and Jacobson do not find the project of mapping emotions with certain core relational themes a pointless endeavour. As D’Arms (2024) stresses: “Concepts and terms such as “offense”, “dangerous”, and “shameful” are, whether tacitly or explicitly, devices for thinking and talking about what it takes for something to be an appropriate object of certain affectively charged motivational states” (p. 631). It is important for us to discuss when emotions are fitting, so the fact that we try to capture an emotion with such interpretive glosses is nevertheless a worthwhile undertaking. They also agree that “there are clearly better and worse interpretations” (D’Arms & Jacobson, 2023, p. 142), but I will get to that in a moment.

So, regardless of whether we take the alethic view which links emotions with core relational themes by saying that each emotion type *represents* a certain core relational theme, or whether we take the interpretive approach according to which mapping emotion types with certain core relational themes is just an *interpretive* undertaking that involves trying to reconstruct emotional valuation in terms of verbal discourse, in both cases, talking about the core relational theme is nevertheless a worthwhile undertaking and should be linked with how we understand the fittingness of emotions. I believe that here also lies the reason why cognitive theories have had such a long reign in moral philosophy and why much of the work on guilt (and other emotions) that has been done through the lens of cognitivism nevertheless proves so insightful. As a discipline, moral philosophy is not particularly concerned with (and does not really have the means to study) the physiological or neurobiological underpinnings of emotion. To come back to Prinz’s analogy of the fuzz buster: a moral philosopher need not care much about the inner structure of the beep. They are likely to be much more interested in what it is that the beep represents or how we can conceptually interpret the beep. In other words, moral philosophers are likely to be more interested in the fittingness of emotion, and the role they play in the narrative flow of human life.<sup>45</sup> And for that purpose, the question of what kind of link exactly it is between the emotion and the core relational theme is not necessarily of the most crucial importance (as long as there is a link and the emotion is matched with the right core relational theme). So, many moral philosophers like Robert Solomon or Martha Nussbaum have construed the link in the simplest cognitivist terms. But much of what they have to say about the fittingness has not really to do with the cognitivist form of their theory but the fact that the content that they attribute to the relevant judgments nevertheless reflects the core relational theme of the relevant emotion type.

Furthermore, regardless of which view we take, methodically, the endeavour of reconstructing the core relational theme is the same. Even though D’Arms and Jacobson emphasise this in the context of the interpretivist approach, I think their idea applies to the alethic view as well: a good formulation of the core relational theme must be grounded in what elicits the emotion and what satisfies it (D’Arms

---

<sup>45</sup> The importance of the role that emotions have in the context of the narrative flow of human life is also emphasised in (Roberts, 2003, p. 3).

& Jacobson, 2023, p. 142). Even for someone who takes the teleosemantic approach, according to which emotions represent core relational themes because they have the function of being elicited by certain stimuli, it is important to look at what it is that actually elicits the emotion (and makes it cease). D'Arms and Jacobson also offer a further test, which again, seems to apply equally well to the representational approach:

“A crucial test of any gloss is whether people with disparate sensibilities can accept it as capturing in a rough-and-ready manner, the way they take the object of emotion—whether or not they deem it fitting. If *dangerous* is a good account of fear’s appraisal, then even a self-aware phobic, who knows the statistics but is nonetheless afraid of flying, will acknowledge that his fear appraises the prospect of his flying as dangerous.” (D’Arms & Jacobson, 2023, p. 142)

Notice that none of this requires that the experience of the emotion must necessarily involve *thinking* of the object of the emotion in terms of the core relational theme of the particular emotion type. What is important is that, upon reflection, they would be bound to agree that the core relational theme captures how they are emotionally attuned towards the object of the emotion.

This is important for our inquiry about guilt. Regardless of which emotion theory we adopt or whether we take the alethic or the interpretive view, if we keep the core relational theme intact and view it exclusively as a matter of moral wrongdoing, then this does not seem to univocally resonate with all people who report guilt feelings for mere harm to others and survivor guilt. Remember the case of Patricia, who got impatient when people tried to comfort her by assuring that the accident was not her fault. What she responded was “Yes, it was an accident [---] but, at the end of the day, I hit him, I took his life”, “I hit him! Why does nobody understand this?” (Gregory, 2017). This response indicates that the idea that she is morally at fault does not seem to resonate with her emotional experience. This is not the response of the self-aware phobic who might say: “I know that flying is safe *but* I just can’t help being afraid”. If she experienced the emotion as a case of recalcitrant guilt, it would be more appropriate to say “I know that it’s not my fault *but* I cannot help feeling guilt”. A response like that implies that the person perceives their emotion as unfitting, as being at odds with their best judgment. Patricia’s response does not indicate that not judging oneself as morally at fault would be at odds with her emotion. She seems to think that her guilt is about something else (about the fact that she hit another person who died) and that there is an important sense in which the emotion gets things right. When she cries, “I hit him! Why does nobody understand this?”, she is saying that it is not her emotion that does not get things right, it is those other people who do not get her emotion right.

My aim in this dissertation is to focus on questions that are guilt-specific, so I will not attempt to resolve the debates between different theories of emotion. I have taken care to show that even if we opted for the feeling or the motivational theory of emotion, this alone would not be sufficient to properly accommodate

the cases of guilt without moral wrongdoing, because the more serious, and the more guilt-specific, debate takes place on the front of the core relational theme. By focusing on the core relational theme, it is my hope that there is nevertheless something illuminating that can be said about guilt that cuts across the boundaries of emotion theories.

### 3. GUILT FOR HARM TO OTHERS

This chapter is dedicated to how we should interpret those cases of guilt that involve merely causal rather than moral responsibility for harm to other people. Those cases were outlined in greater detail in section 2.2.3, and spanned across a wide range of scenarios that included car crashes, being a reluctant agent of another person's heartbreak, being the remote cause of Holocaust (the Jewish boy who did better than Hitler), having a rescuer die in an attempt to save you, being used as an instrument of evil, taking space in an overloaded wagon, and miscarriage. Here, the focus is on how to make sense of those cases as instances of guilt. In particular, I am interested in how those cases have pushed the theorisation about guilt towards an interpersonal approach that eliminates the moral transgression component and puts the focus on simple harm to others. First, I will give an overview of the interpersonal account of guilt, as defended by Roy Baumeister and his colleagues (1994), and explain what is at stake if we interpret guilt in social rather than strictly moral terms. As I will explain, it sheds doubt on the idea that guilt is a moral emotion in a traditional sense—as an emotion, the explanation of which involves a moral term. On the background of critical concerns about the overmoralisation of emotions (D'Arms & Jacobson, 2000; Prinz, 2009), this may indeed seem like a welcome development. Most emotion types do not have a moral shape, so why should we think that guilt has? Yet if there are any strictly moral emotions at all, then guilt would seem like the most obvious candidate. So, in that sense, the case of guilt seems like one of the most decisive cases with regard to whether we can speak of strictly moral emotions at all. As I will go on to argue in the second part of this chapter, cases of merely causal harm should not make us give up on the idea that guilt is a moral emotion. Throwing morality out of guilt just because it does not always involve moral wrongdoing is like throwing the baby away with the bathwater. Even though guilt need not involve perceiving one's *deed* as a morally *wrong* one, a closer look at the cases that the proponents of the interpersonal approach refer to reveals that those cases are nevertheless related to perceived moral *badness* of the *consequences*. First, I will suggest that the chief good that guilt for harm to others is linked with—the well-being of other people—is nevertheless the same good that the subject views as a basis of their deontic duties as well, and as such, we can rightfully call it a moral good. Secondly, there is also a second layer of moral badness that has often gone unnoticed—cases of harm to others are concerned with *undeserved* harm. As such, I will argue that cases that may, at first sight, seem like non-moral matters, actually contain multiple layers of moral concerns.

## 3.1. The interpersonal account and the non-moralisation of guilt

### 3.1.1. The interpersonal account of guilt

According to the interpersonal account of guilt, guilt is not a matter of moral norm violations but should instead be understood in terms of causing harm to other people. This account of guilt was originally proposed by Roy Baumeister and his colleagues in their landmark article “Guilt: An Interpersonal Approach” (1994). The authors explicitly contrast their approach with the so-called intrapsychic account of guilt, which is basically (at least in its best form)<sup>46</sup> the same as the traditional account of guilt that I have been talking about in this dissertation. Yet Baumeister and colleagues insist that guilt is not just something that happens “inside people”—people evaluating themselves against abstract moral norms –, it is also something that happens “between people”, which highlights the fact that guilt involves sensitivity to the harm that happens out there in the world. (Baumeister et al., 1994, p. 243) This clash between the intrapsychic and interpersonal approaches has been one of the most important divisions within the psychological theorisation of guilt for a long time (Carni et al., 2013), and has inspired many further debates, some of which, for example, concern the idea that perhaps both approaches capture a somewhat different type of guilt (Gangemi & Mancini, 2011; Mancini & Gangemi, 2021).

From a conceptual aspect, there are two types of questions that we might ask about the interpersonal account of guilt: is perceived harm to others *necessary* for all instances of guilt and is harm to others *sufficient* to capture what guilt is about? Even though this will not be my main focus, let me briefly touch upon the first question, as this will help us put the interpersonal approach into a wider perspective. At least one reason why it might be doubted whether interpersonal harm is necessary for guilt is that certain violations of religious morality like masturbation and not honouring the Sabbath (Mancini & Gangemi, 2021, p. 7) seem to pose a counterexample. It is not clear at all that someone who felt guilt about not having honoured the Sabbath and working on Sunday would first and foremost be concerned with the fact that they have harmed anyone. The problem seems to lie in the idea that they have not paid sufficient honours to God, but God in its perfection does not seem to be correctly described as harmed by it. Neither does guilt about masturbation seem to lie in concern for harming anyone. Of course, above everything else, one might also view masturbation as harmful for oneself, but the reason why they consider it immoral is more likely to lie in the idea that it goes against the natural God-set order. Both of those examples seem to fare better with the traditional, intrapsychic approach, according to which guilt is just

---

<sup>46</sup> Sometimes, intrapsychic approach is also associated with self-punishment (Baumeister et al., 1994, p. 244), but I will focus on the intrapsychic approach in the form of the traditional theory of guilt in terms of perceived moral wrongdoing because I think that this represents the intrapsychic account in its strongest form.

a matter of moral norm violations.<sup>47</sup> But this is a debate that I will not enter in this chapter. It is not my intent to give a full analysis of all the aspects of the interpersonal account of guilt. Even if the interpersonal account of guilt does not capture all instances of guilt, it may nevertheless do an excellent job in capturing an important type of guilt (as also suggested by Mancini and Gangemi 2021). Since the aim of this dissertation is to make sense of the practice of extending the word ‘guilt’ to cases without perceived wrongdoing, that is, cases of merely causal harm and survivor guilt, I am first and foremost interested in how and to what extent the interpersonal account of guilt can help us illuminate precisely those cases—cases that were motivating the shift towards the interpersonal account in the first place. In that sense, I am interested in how well the interpersonal account fares with regard to the cases that it explicitly purports to explain. As it will be seen in Chapter 4, in the case of survivor guilt, I do not think that it is even necessary to perceive oneself as having caused the harm that befell others. This chapter, however, is a deep dive into guilt for merely causal harm. In those cases, concern for the harm that one has caused is already built into the scenarios as an integral element. So, the main question that I will be concerned with in this chapter as a whole is whether perceived harm to others is indeed sufficient to explain what those instances of guilt are about, and eventually I aim to show that those cases are not really about harm simpliciter but involve a moral element as well.

But before I move on to my criticism (in 3.2), I would first like to highlight some of the advantages of the interpersonal account of guilt. Indeed, one of its greatest advantages is that, at least compared to the traditional account, it does an especially good job explaining cases of merely causal harm. Those involve, for example, the guilt felt for causing a car accident where someone got killed even if it is clear that everything happened so quickly that there is nothing that the driver could have done (Zhao, 2020) or being the reluctant agent of someone else’s heartbreak (Baumeister & Wotman, 1994). But it is important to note that the interpersonal account is not just meant to accommodate those cases, but the majority of the more traditional cases as well—it reinterprets them as first and foremost matters of causing harm. Sometimes, causing harm may indeed come in the form of violating a moral norm, but the more essential element is nevertheless the harm itself rather than the fact that one has violated a norm. What is important is the empathetic identification with the victim’s suffering. Does this mean that the account may have difficulties with explaining those instances of the traditional norm-violation guilt that do not involve harm (like the religious ones that I previously listed)? Perhaps so. But, on the other hand, it opens the road to a completely different way of thinking about guilt, where the conceptual dominant is something other than traditional moral norm violations. Even though I will not eventually agree with all the details of the account, my thinking about guilt has been largely indebted to the interpersonal account, and I also believe that, in great

---

<sup>47</sup> For more counterexamples, see also Jesse Prinz’s discussion of victimless crimes (Prinz, 2011b, p. 214, 2011a, p. 218).

part, the interpersonal account takes things in the right direction by putting the focus on the more outward- rather than inward-looking elements of guilt. Interestingly, it was just a year before the seminal article by Baumeister and his colleagues that Bernard Williams highlighted in his “Shame and Necessity” that the modern approach to guilt that puts a great emphasis on the moral culpability of agents runs the risk of taking guilt to an overly self-indulgent and narcissistic extreme where it becomes a matter of mere self-flagellation, whereas it would be a welcome development to focus more on the suffering of the victims rather than the self (B. Williams, 1993, pp. 219–223)—as in case of agent-regret, which he considers as “psychologically and structurally a manifestation of guilt” (Williams 1993, 93, see also Zhao 2020, 298–9). The interpersonal account does precisely that.<sup>48</sup>

What is especially interesting about the specifics of the interpersonal account, though, is that it altogether loses the reference to morality, so that guilt ceases to be a moral emotion in the strict sense of the term. What is a moral emotion in this strict sense that philosophers have long used the term? To call an emotion type moral means that what we have been calling the core relational theme of the emotion, i.e. that which the emotion type, in general, is about and how it represents its object, involves a reference to a moral concept.<sup>49</sup> Fear, for example, is not a moral emotion because its core relational theme concerns danger and dangerousness is not an inherently moral matter. Guilt, at least according to the traditional understanding, however, is a paradigmatic example of a moral emotion because its core relational theme is supposed to concern the subject having done something morally wrong, which certainly is a moral matter. But once we adopt the interpersonal account of guilt, which locates the core relational theme in causing harm rather than violating moral norms, the direct reference to morality disappears and guilt ceases to be a moral emotion in a strict sense.

Of course, this does not mean that we cannot talk about guilt as a moral emotion at all. Even if one adopted the interpersonal view, it may nevertheless be

---

<sup>48</sup> Of course, there is a sense in which Williams and Baumeister are engaged in different projects. Williams is in the business of critiquing the contemporary morality system and the kind of norm-violation guilt that goes along with it, Baumeister and colleagues just try to capture how people actually experience guilt. Williams does not deny that the norm-violation account of guilt captures how people usually feel guilt, he just points out that as a society we might be better off focusing on a different kind of guilt that is more concerned with the suffering of the victims. So, for Williams, an account that says that guilt is about harm to others in the first place might be a laudable step in terms of conceptual engineering.

<sup>49</sup> Different philosophers may put this in slightly different terms. John Rawls distinguishes moral attitudes from natural ones by requiring that “the person’s explanation of his experience invokes a moral concept and its associated principles. His account of his feeling makes reference to an acknowledged right or wrong.” (Rawls, 2005, p. 481) Aaron Ben-Ze’ev, who speaks in terms of core evaluative concern, says: “If this intrinsically normative concern refers to moral issues, then the emotion can be regarded a moral emotion” (Ben-Ze’ev, 2002). Others may prefer to speak in terms of the (formal) object of the emotion as instantiating some moral matter (Mulligan 2008, see also the first definition listed in Cova *et al.* 2015). Yet others may say that “moral emotions present their object in the light of such moral concepts such as desert, fault, and responsibility” (D’Arms & Jacobson, 2000, p. 87). But the idea is the same.

argued that guilt remains one of the most prototypical moral emotions in a looser sense. According to Jonathan Haidt (2003), being a moral emotion is a matter of degree, and depends on how many emotion tokens of the particular emotion type are either elicited by morally relevant situations or contribute to the moral functioning of the world. So, if sufficiently many instances of guilt are still elicited by moral wrongdoing and/or contribute to morally good ends (by motivating the person to make amends and change their ways), then guilt will nevertheless be a moral emotion in this looser sense to a relatively high degree—perhaps even to the highest degree compared to all the other emotion types. So, even if it was the case that guilt is about harm simpliciter, this does not necessarily mean that we cannot talk about guilt as a moral emotion at all. And many may indeed be satisfied with this and find it more appropriate to talk about moral emotions in this looser sense.

However, the claim that guilt is a moral emotion in this looser sense is a fairly trivial claim. What is more interesting, at least from a philosophical point of view, is whether we can talk about guilt as a moral emotion in the traditional sense. Guilt has been a stock example of a moral emotion in this stricter sense, and if not even guilt lives up to the stricter standard of being a moral emotion, then it may well be doubted if any emotion can. Given this focus, henceforth, I will only be talking about moral emotions in the strict sense and will drop the somewhat cumbersome qualifier “in the strict sense”.

### 3.1.2. The nonmoralisation of emotions

The idea that guilt is not a moral emotion makes the interpersonal account of guilt an especially interesting and welcome development in the background of the wider philosophical trend of non-moralising the shapes of emotion. Those two developments have not usually been viewed in conjunction, but understanding this wider movement toward non-moralising the shapes of emotion is important because it can be taken as providing further motivation to adopt a non-moral account of guilt like the one proposed by the interpersonal approach. So, let us review the movement toward non-moralisation of emotion shapes in greater detail. First, I will focus on reasons pertaining to our own nature, i.e. various psychological biases, that should make us suspicious about the all too human tendency to overmoralise the shapes of emotions. Most notably, I will look at what Justin D’Arms and Daniel Jacobson (2000) have thematised as the moralistic fallacy, but I will also complement their analysis with some further psychological biases that may play a role. Secondly, following Jesse Prinz (2008, 2009), I will also review the reasons pertaining to the nature of emotions themselves that should make us wary about the idea that emotions have a neatly moral shape.

Let us begin with the suspicious psychological tendencies that may push people towards the overmoralisation of emotions. An important landmark within the trend of non-moralising the shapes of emotions is an article entitled “The Moralistic Fallacy” (2000) by Justin D’Arms and Daniel Jacobson. The authors suggest that both ordinary thought and philosophical theorisation often suffer

from the so-called moralistic fallacy, which consists of inferring “from the claim that it would be wrong or vicious to feel an emotion, that it is therefore unfitting” (D’Arms & Jacobson, 2000, p. 69). To understand what they mean, let us first go over the notion of fittingness. Fittingness, or emotional correctness, concerns whether the emotion token correctly presents the particular object of the emotion as having certain evaluative features (p. 66). In other words, we might also say that it concerns whether the scenario at hand instantiates the core relational theme of the emotion type. For example, fear would be considered fitting if the particular object of fear really is dangerous. It is in this sense that the basic (evaluative) shape of the emotion fits the particular scenario. So, what they are saying when they talk about the moralistic fallacy is that there is a common erroneous tendency to jump from thinking that some emotion token would be morally inappropriate to the idea that the emotion must be presenting the world in an incorrect way. Their main examples of emotions that have fallen prey to the moralistic fallacy are envy and amusement. In a moralistic vein, it may be suggested that if a joke is offensive and, therefore, morally inappropriate, then it cannot be funny. Similarly, it might be thought that since it would be morally inappropriate to envy your recently tenured colleague who truly deserved the promotion, then such promotion is not enviable in the sense of fittingness. This is possible if we build moral appropriateness conditions into the fittingness conditions of envy and suppose that envy presents one’s rival as *undeservedly* possessing some good that the subject themselves lack. Then, it would indeed be the case, that it would be unfitting to envy your colleague for a deserved promotion. Yet it seems farfetched to say that someone in the grip of those emotions must necessarily perceive the world in moral terms—that they must see the joke as morally appropriate and the promotion as undeserved. To jump from the mere fact that a particular emotion token would be morally inappropriate to the idea that it must, therefore, be unfitting is a mistake because moral appropriateness and fit are two separate matters. Yet D’Arms and Jacobson also find it natural that people are prone to such fallacious inferences because “it is easy to be uncomfortable with the idea that a morally dubious emotion might nevertheless be fitting” (p. 75). This reflects a more general psychological tendency that they call moralism—“the imperialistic tendency of moral considerations to take over the entirety of evaluative space” (p. 75).

D’Arms and Jacobson do not deny that there may be emotions that have a moral shape. In fact, they suggest guilt and anger as candidates for emotions that may indeed present their objects in light of moral concepts like fault and responsibility (D’Arms & Jacobson, 2000, p. 87).<sup>50</sup> What they are explicitly against is the fallacious inference from moral inappropriateness to unfittingness. But

---

<sup>50</sup> In other works, they also suggest that resentment as a cognitive sharpening of anger that involves a sense of unfairness may have a moral shape (D’Arms & Jacobson, 2003, p. 143). But then, we would not be dealing with a pan-cultural natural emotion anymore, like amusement, anger, contempt, disgust, embarrassment, envy, fear, guilt, jealousy, joy, pity, pride, shame and sorrow.

overall, their article has the effect of making the reader suspicious of moralising the shapes of emotions as a more general phenomenon—because it hints that one may have fallen prey to the moralistic fallacy or moralism in a more general sense. So, even in the case of guilt, we may wonder (even though the authors themselves do not) whether it is perhaps just because of the idea that guilt feelings in the absence of wrongdoing are morally inappropriate (at least in the sense of not being required) that the proponents of the traditional account of guilt have been prone to think that such guilt feelings cannot be fitting.

But there is more to be said against the overmoralisation of the shapes of emotions. In particular, I am interested in the more general tendency that D’Arms and Jacobson call moralism—“the imperialistic tendency of moral considerations to take over the entirety of evaluative space” (D’Arms & Jacobson, 2000, p. 75). While D’Arms and Jacobson do mention this tendency, this is not their main focus. Their main concern is unmasking an erroneous way of thinking that involves jumping from moral inappropriateness to unfittingness. Perhaps it is understandable that they do not engage in too much psychological speculation about moralism in general but keep their focus on the wrong ways of argumentation that have been evident in philosophical writing. But it is nevertheless interesting to ask further what might be behind the general tendency toward moralism. One option is that we are dealing with sincere overdue concern with morality. This may be because people sincerely care about morality (like supposedly the people who think that if a joke is morally inappropriate, then it cannot be appropriate in any sense of the term). In a more professional context like moral philosophy, it may also be combined with the fact that moral philosophers are so used to thinking about the world in moral terms that they tend to see all matters through the prism of moral concepts. All this reflects an undue concern for morality, which leaves no room for admitting that emotions may also have other criteria of appropriateness besides the moral one.

But in other cases, the tendency to build moral concepts into our definitions of emotions may also be due to our tendency to employ morality for self-justificatory or otherwise self-serving purposes. Even though this idea remains largely unthematized in the contemporary talk about moral emotions, the idea is itself not new—it was already in the 19<sup>th</sup> century that Friedrich Nietzsche drew our attention to the idea that taking the moral high ground can be just a petty and self-deceptive way to boost one’s self-esteem (Nietzsche, 2007).<sup>51</sup> Even though we might not agree with Nietzsche’s blanket rejection of morality, it is undoubtedly the case that sometimes our talk about morality can indeed be employed for self-serving purposes. I will illustrate this with the example of anger, envy and *schadenfreude*, which seem to run an especially high risk of being combined with a morally

---

<sup>51</sup> In the first tractatus of “On the Genealogy of Morality” (2007), Nietzsche suggests that our contemporary morality that is heavily influenced by Christianity arises from the *ressentiment* of the weak against the strong. Since the weak do not possess strength, they start to think that strength is bad, just like a fox that does not reach for grapes can come to think that grapes are sour and not worth reaching for.

problematic kind of self-righteousness,<sup>52</sup> but I think that the same ideas can apply to many other emotions as well. When we are the ones feeling angry towards someone, it is beneficial for us to think that our anger captures a moral offense because this would bolster our claim to insist that the other person change their ways—it helps us make the claim that they have not just frustrated our egoistic goals but committed a moral offense. Thus, we can easily come to think that anger itself must be intrinsically concerned with morality, even though it is we ourselves who add the moral story on top of anger without much sincere concern for morality. Also, it is beneficial for us to think of anger in moral terms when we are on the receiving end of anger. It gives us a better chance to deflect other people's anger. We can provide excuses to show that we were not morally at fault and if anger is exclusively about moral offenses, then showing that we were not at fault should be sufficient to make the claim (both to the other person and also third parties who might come to our defence) that the other person's anger is totally misplaced. In contrast, if anger was simply about goal frustration, it would be much harder to show that we have not indeed frustrated other people's goals and claim that the other person has no reason whatsoever to be angry at us.

Emotions that are concerned with comparative status, may also easily tempt us to frame them in moral terms. I am talking about envy and *schadenfreude*, which arise from concern for comparative status—in case of envy, we are frustrated about our rival faring better than us, while in case of *schadenfreude*, we are happy to see our rival suffer a misfortune.<sup>53</sup> When we are the ones who are envious, appealing to the idea that the other person does not really deserve their good fortune helps to bolster our self-esteem—we can still think of ourselves as being more worthy. On the reverse, if someone envies us, tying envy with morality gives us a better chance to deflect their envy as thoroughly misplaced if we can show that we are indeed deserving. Moving on to *schadenfreude*, when we feel *schadenfreude* it can certainly make us feel more righteous if we think that our rival's misfortune is deserved. But does it really make sense to say that *schadenfreude* is, at its core, a moral emotion that reflects our concern for justice (as defended, for example, by Portmann 2000)? I do not think so. It seems to me that if there is sufficient rivalry between two people, and one really wants to enjoy a higher status than their rival, then the mere fact of misfortune, be it perceived as deserved or undeserved, can be enough for the other one to rejoice gleefully. Similarly, when we think about other people feeling *schadenfreude* towards us, it is for self-serving purposes that we may build desert into the formula—because if we assume that *schadenfreude* presents us as suffering a deserved misfortune, then this makes room for us to argue that the suffering is not deserved and the other person's *schadenfreude* is in every sense inappropriate. But none of that

---

<sup>52</sup> This can also be related to the vice of pride, in particular the kind of moral superiority that some people may be prone to assume. See more about the vice of pride and the dangers of moral superiority from (Roberts, 2009).

<sup>53</sup> See more about the connections between the two emotions with regard to personal comparative concern from (Ben-Ze'ev, 2014).

reflects true concern for desert, and the root of the emotion seems to be elsewhere—in a more egoistic desire for advancements in comparative status.

The examples of anger, envy and schadenfreude demonstrate why our tendency to overmoralise the shapes of emotions might be not only due to sincere concern for morality but may also reflect a more egoistic concern—to strengthen our claim for being in the right, to deflect other people’s negative responses, to boost our own self-esteem. This does not only happen at the individual level but may sometimes take place at the species level as well. For example, it has been suggested that evolutionary ethicists, ethologists, developmental psychologists, social neuroscientists and cultural anthropologists may fall prey to an “almost irresistible” tendency to postulate an innate moral faculty (and thus also the possibility of moral emotions) because it would enable us to view ourselves nobler as a species (Prinz, 2009).<sup>54</sup>

Could it be that such egoistic concerns have played a role in the overmoralisation of guilt? Since the only moralised account of guilt that we have discussed so far is the traditional one that views guilt through the lens of moral wrongdoing, let us discuss on the example of this account. At first glance, it may be hard to believe that the painful realisation that one has done something wrong could serve an egoistic purpose. Differently from anger, envy, and schadenfreude, guilt does not involve any complaint or negative judgment about the other person. It is a self-critical emotion that represents the subject as being at fault. This does not seem like a particularly fruitful way to boost one’s self-esteem. But perhaps sometimes it can. For survivors of extreme violence, it can be hard to accept that they were in a totally helpless state, robbed of any kind of control over what happened to them—as a mere thing pushed around by foreign forces. As suggested by Amber Griffioen (2014), in those cases, guilt can help the traumatised person to rebuild their sense of self because it allows the traumatised person to retrospectively maintain a sense of agency, and “*it is better to be a guilty agent than no agent at all*” (p. 51). She also refers to Susan Brison, a philosopher and rape survivor, who reports based on personal experience that “it can be less painful to believe that you did something blameworthy than it is to think that you live in a world where you can be attacked at any time, in any place, simply because you are a woman” (Brison, 2002, p. 13; Griffioen, 2014, pp. 51–52). So, it does indeed seem that, at least in some cases, tying guilt with moral wrongdoing can offer a boost to one’s self-esteem. But the cases where one has fallen victim to extreme violence are nevertheless quite extreme cases. Furthermore, those cases are explicitly discussed as instances of irrational guilt. Does the fact that guilt can offer a boost of self-esteem after experiencing such a severe trauma mean that guilt is, in general, a self-boosting emotion? Does it offer a self-boost after one has lied, stolen, or committed a murder? As already said, for the most part, guilt seems to

---

<sup>54</sup> Prinz himself thinks that such a stance “oversells human decency and undersells human potential” (Prinz, 2008, p. 368) because if morality, rather than being innate, would instead be a flexible tool, something that we construct, then this would mean that we can also reshape our moral systems towards the better (Prinz, 2008, p. 406).

offer self-criticism rather than self-boost. But even though one's self-esteem might not be boosted compared to their pre-crime self-esteem, there is still a sense in which it is better to think of oneself as an agent in those more regular instances as well. As put by Maria Miceli and Cristiano Castelfranchi in their comparison of shame and guilt:

Both shame and guilt are “self-critical” emotions. However, self-criticism may take different self-evaluative forms: on the one hand, people may view themselves as ugly, stupid, handicapped, or morally defective—in a word, *lacking* (in physical attractiveness, intelligence, skills, moral worth, and so on); on the other hand, they may view themselves as wicked, unjust, sinful—that is, *endowed with the power to violate norms and thwart others' goals, and willing (or inclined) to do so*. (Miceli & Castelfranchi, 2018, p. 711)

Thinking of oneself as an agent who has control over their actions means thinking of oneself as *powerful*. Regardless of which instance of guilt we are talking about, thinking of oneself as powerful always adds a silver lining to an otherwise painful emotion. This self-boost may be limited in the sense that it does not erase the self-critical part of the emotion, but even a limited self-boost is still a self-boost. I am not arguing that the main problem with the traditional account of guilt is necessarily the fact that it involves the overmoralisation of guilt for self-interested reasons. But, looking at the matter on the background of general concerns about the overmoralisation of emotion types, it is fair to note that the possibility that there may be a self-interested incentive to overmoralise guilt is still there.

All these psychological tendencies provide us with a genealogical<sup>55</sup> reason to be cautious about the moralisation of emotion shapes. They draw our attention to the fact that, oftentimes at least, the moralisation of emotion types has suspicious origins—it derives from our fallacious psychological tendencies. As such, genealogy defamiliarises—it reminds us that the fact that some emotion type seems moral may simply be an illusion due to our own biases.<sup>56</sup> This, however, as D'Arms and Jacobson (2000), too, are well aware of when they talk about the moralistic fallacy, does not necessarily mean that our fallacious reasoning must have led us to the wrong conclusion—because the conclusion may be right even if it is reached to in an incorrect way. In that sense, the talk about the moralistic fallacy and other moralistic tendencies tells us more about ourselves than about the emotions—the reason for being cautious about moralising the shapes of emotion derives from our own fallacious tendencies rather than from some general fact about emotions. In what follows, I would like to complement the previous talk by

---

<sup>55</sup> I am indeed referring to the sense of genealogy that was known to be used by Friedrich Nietzsche (2007) but has later been employed by many other authors as well, including Martha Nussbaum (2016, p. 58).

<sup>56</sup> I am here paraphrasing the idea expressed by Nussbaum (2016), when she highlights the advantages of genealogy: “History defamiliarizes. It reminds us that things could be, and indeed sometimes were, otherwise” (p. 58). However, while she is talking about historical genealogies, I am applying the same idea in the context of psychological genealogy.

highlighting some general facts about emotions themselves that might make us suspicious of the idea that emotions come in moral shapes.

We find this kind of more ambitious argument from Jesse Prinz, who is much more radical than D'Arms and Jacobson. Indeed, Prinz goes as far as to say that if the term "moral emotions" is taken to designate a proprietary class, it should be considered "something of a misnomer" (Prinz, 2009, p. 519). If I have understood him correctly, his overarching argument relies on the idea that basic emotions do not have an exclusively moral shape, and all other emotions are just blends of basic emotions, and for this reason, can also crop up in nonmoral contexts. To some extent, the way he sets his argument up seems case-based—he simply goes over a number of emotions that are often associated with morality (anger, disgust, contempt, disappointment, guilt, shame and fellow feelings) and shows for each that it is either a basic emotion for which it is quite obvious that it serves functions outside the moral domain, or it derives from other, more basic emotions that are not distinctively moral, and can thus pop up in non-moral contexts as well. For example, even though anger is often treated as a matter of moral offence, it seems more correct to say that it is just a response to goal frustration in general because we can also be angered by irritating sounds, difficult and recalcitrant tasks, or when someone glares at us. At least this is how anger appears in its evolutionarily ancient form. Disgust, which is often associated with religious morality, is not exclusively tied to morality either—it can be felt in response to contaminants in general, and even the so-called moral disgust may simply be physical disgust that is felt in response to moral norm violations that concern the purity of the body. Contempt, he considers as a blend of anger and disgust, and perhaps feelings of dominance, but in any case, no one seems to doubt that it can arise in nonmoral contexts—the rich can harbour contempt for the poor, and vice versa. (Prinz, 2009, p. 535) It is beyond the scope of this section to review Prinz's case-based argument in full and go over what he says about all the emotions that have been considered morally relevant but I have knowingly chosen to present precisely those examples because these are the emotions that are part of the famous CAD-hypothesis (Rozin et al., 1999) that maps Richard A. Shweder's "Big Three" of ethical discourse (Shweder et al., 1997)—the ethics of Community, Autonomy, and Divinity—respectively with the emotions of contempt, anger, and disgust.<sup>57</sup> Since contempt, anger and disgust do not seem to have a moral shape, this goes in some way of showing that, instead of starting out with emotions with a uniquely moral shape, societies simply make use of natural emotions (that do not have a uniquely moral shape) to moralise certain types of behaviours.

---

<sup>57</sup> I am aware that these days it is more common talk about the five domains of morality as specified by Jonathan Haidt and Craig Joseph (2004). The reason why I am not here talking about their domains is that they have no pretensions of mapping the domains neatly with specific kinds of emotions. So, they would not provide us no reason to think that there are neatly moral emotions. Furthermore, when proposing his own definition of moral emotions, which I already discussed earlier, Haidt (2003) is clear that being a moral emotion is a matter of degree, so to the extent that he does speak of moral emotions, he does not rely on the idea that there are moral emotion in the more traditional sense that I am interested in.

This is in line with Prinz's general position, according to which morality is not innate but a by-product of capacities that have evolved for other purposes. We have to be a little bit careful here not to get the order of argumentation wrong. To some extent, his general case against morality as innate rests on the very idea that there are no exclusively moral emotions rather than providing us with a reason to think that there are no moral emotions. It is for this reason that he simply presents it as a potential implication of his case against moral emotions that morality emerges out of more general psychological resources (Prinz, 2009, p. 536). That being said, he also has many other reasons for thinking that morality is not innate, and I think that in a more holistic sense, those other reasons could be taken as providing indirect support for the idea that emotions, too, do not have an exclusively moral shape. What are those other reasons? As Prinz highlights, solid evidence for the idea that morality, in general, is something innate seems scarce. For ethical reasons, we do not have studies of people who grow up without any moral training, so that their morally sensitive behaviour could function as evidence for morality being something innate, and the anecdotes about children who have grown up alone in the woods do not seem to tell a particularly convincing story about their moral sense (Prinz, 2008, pp. 393–394). Also, even if it was the case that we could detect some moral universals across different cultures (something which Prinz is actually sceptical about), it is extremely difficult to prove that this is necessarily due to the fact that we have some innate moral faculties rather than the fact that, by making use of more general psychological resources, societies just develop moral rules due to situational factors (p. 405). What kind of situational factors? For example, since, due to their ability to reason, people exhibit greater potential for bad behaviour than other animals (p. 405), this calls for a countermeasure if a society is to thrive as a whole. There has also been the added pressure of having to cope in societies with an ever-growing number of people, where the strong bonds that we have towards our kin are not sufficient to grant social harmony. Given those situational pressures, it is no wonder that cultures start to inculcate moral rules and values. They develop morality as a “tool” to solve those situational pressures (p. 406). In that sense, Prinz suggests that morality is similar to religion, which, too, has appeared everywhere, but as a cultural construct rather than the result of an innate religious sense (p. 404). And last but not least, if morality was truly hard-wired, then it would make sense to expect that it emerged through incremental biological evolution and other species, too, have rudimentary versions of those moral traits that we attribute to humans. Unfortunately, it is beyond the scope of this chapter to review Prinz's reinterpretations of those cases that have given rise to the idea that nonhuman animals exhibit a moral sense,<sup>58</sup> but the general idea is that many of those cases can be explained without reference to morality, and the fact that nonhuman animals experience vicarious distress and may have prosocial tendencies is not sufficient

---

<sup>58</sup> In particular, he is addressing studies about animals who stopped pressing a lever to release food if, while pressing the lever, they saw another specimen of their species (Church, 1959; Masserman et al., 1964; Watanabe & Ono, 1986; Brosnan & De Waal, 2003).

to talk about distinctly moral traits but is just an example of general psychological capacities, which can later (in humans) be employed in the service of morality as a culturally constructed tool (pp. 397–402).

Even if Prinz does not provide us with a completely airtight argument against the possibility of moral emotions, the considerations that he draws attention to nevertheless place a heavy burden on anyone who wants to defend the existence of moral emotions. In particular, I am sympathetic towards Prinz's idea that morality is not innate and that those emotions that could be linked with the idea of morality as innate, that is, basic emotions like fear, disgust and anger which have a long evolutionary history and analogues in non-human animals do not come in a moral shape. This is probably due to the fact that those truly long-term evolutionary challenges in response to which we have developed those emotions do not come in a highly moralised form. Survival of the fittest does not inherently care about morality, and it would indeed be a great surprise if our most basic emotions turned out to have a uniquely moral shape. So, I am willing to grant that, when it comes to basic emotions, Prinz makes a relatively strong case against moral emotions.

However, I do not think that this part of his argument really applies to guilt because, compared to other emotions out there, guilt seems like an especially weak candidate for a basic emotion.<sup>59</sup> It does not even seem culturally universal—many cultures seem to be regulated by shame rather than guilt (e.g. the ancient Greeks and the contemporary Japanese).<sup>60</sup> Furthermore, even if we can detect an evolutionary challenge to which guilt might provide a solution—e.g. the sustenance of cooperative arrangements in the face of opportunities to defect, or the appeasement of one's peers whom one has harmed—it seems that we already have evolutionarily more primal emotions like fear of punishment and shame that could have already gone a relatively long way in helping us cope with those challenges (Kurth, forthcoming). So, from an evolutionary perspective, it would have been redundant for us to develop a completely new basic emotion (compared to guilt as a culturally developed tool to simply further reinforce cooperation in societies with an ever-growing number).

So, when it comes to guilt, we must look at the addendum of Prinz's overall argument: the idea that all other, non-basic emotions, are just blends of basic emotions, and for this reason can crop up in non-moral contexts (Prinz, 2009, p. 535). The way he applies this idea to guilt is the following:

---

<sup>59</sup> While there are some attempts to defend guilt as a natural kind, for example (Ramsey & Deem, 2022), those attempts are relatively rare. This is why I say that, *compared* to other emotions, the case for guilt as natural kind is weak. The considerations that speak against guilt as a natural kind are listed in what follows.

<sup>60</sup> The distinction between guilt cultures and shame cultures was first introduced by Ruth Benedict in her "The Chrysanthemum and the Sword: Patterns of Japanese Culture" (1946).

Guilt may be linked to sadness. We feel sad about harming people in our in-group because doing so threatens attachment relationships. Guilt may also occasionally encompass anxiety because loss of attachment and punishment often follow when we cause harm. So guilt may be an anxious sadness elicited by the harms we have caused (Prinz, 2004). If guilt and shame derive from non-moral sources, we might expect them to crop up in non-moral contexts from time to time, and indeed they do. Survivor guilt occurs when a person is among the few to survive a calamity that killed members of the in-group. (Prinz, 2009, p. 535)

He is suggesting that guilt is a blend of sadness about the harm that we have caused combined with anxiety about the loss of attachment and punishment that may follow, and that this kind of blend can also crop up in survival situations, which are allegedly nonmoral matters. Let it be said that, for reasons explained in Chapter 4, I do not really find it particularly promising to explain guilt in terms of anxiety, and neither do I think that survivor guilt, because of its connections with injustice, is a particularly good example of a nonmoral form of guilt. But for my current argument, none of this matters. What matters is that guilt is a blend of other, more basic emotions. In that sense, Prinz's account of guilt is reminiscent of Ortony's (1987) idea that guilt should not be considered as an independent emotion on its own but is more akin to feeling betrayed or abandoned, which can both be construed as a whole complex of emotions that the relevant type of situation normally gives rise to. Yet if the mixture of emotions that is associated with guilt is sufficiently distinct, so that the experience of this unique blend has come to represent to the subject (through cultural inculcation) the idea that something has gone morally astray, it is not clear to me why we should think that the fact that guilt may, from time-to-time, crop up in non-moral contexts mean that it is therefore not a moral emotion. Ultimately, what is important is how the subject construes the situation. If the subject nevertheless perceives that situation as a morally relevant one, then the emotion remains a moral one—even if, from a more objective point of view, such construal is incorrect (as we will be assuming for the sake of the argument). Even though emotions can misfire, this does not change the fact that their core relational theme may nevertheless involve a reference to morality. The beep of the fuzzi buster may misfire as well, but that does not change the fact that it nevertheless represents the presence of police radars to the subject.

It seems that what Prinz would need to show in order to make the case that guilt is not a moral emotion is that there are cases where guilt crops up in a perfectly fitting (rather than unfitting and recalcitrant) form, yet does not construe the situation in moral terms. As such, it seems that the argumentative force of Prinz's argument does not really come from the idea that guilt may also crop up in the allegedly nonmoral context of survivorship, but from his initial definition of guilt as a blend of “anxious sadness elicited by the harms we have caused” (2009, p. 535). As such, he is already assuming something (if we disregard the anxiety part) akin to the interpersonal view that reduces the core relational theme of guilt to harm to others. In that sense, the point he wants to prove is already built into his assumptions. He offers us no further argument in defence of the non-

moralisation of guilt than the one we might already find from the interpersonal account itself—the idea that we might be able to construe the core relational theme of guilt in non-moral terms. Perhaps we should not hold it against him. Given both the general problems with the overmoralisation of emotion types and the more specific problems related to the traditional account of guilt in terms of moral wrongdoing being overly narrow to account for guilt for harm to others and survivor guilt, what is needed is precisely an alternative characterisation of the core relational theme of guilt, and he does indeed seem to point towards one that concerns harm to others (see also Prinz and Nichols 2010). Yet if his final argument is ultimately case-based, it must also be possible to contest it in a case-based manner, which is precisely what I will do in the next section.

In this section, I have argued that one of the advantages of the interpersonal account of guilt is that it resonates well with the wider philosophical trend towards the non-moralisation of emotion types. I hope to have shown that concerns about the overmoralisation of emotions that arise from both our own psychological biases and the nature of emotions themselves are indeed serious concerns, and we should not easy-handedly deem an emotion a moral one. Even though the considerations discussed in this section do not provide a completely air-tight case against the possibility of some rare specimens of moral emotions, taken in conjunction, they do place a heavy burden of proof on those who want to defend the existence of moral emotions. On the background of those concerns, it would indeed be an intriguing matter to find an emotion type that we could rightfully call a moral emotion. Is the tradition of talking about moral emotions in a traditional sense just a conservative artefact that should remain in the past, or can it be that we can nevertheless reasonably talk about moral emotions in this traditional sense? Given that guilt is considered as one of the most paradigmatic moral emotions of all, the case of guilt seems to be of decisive importance. If not even guilt lives up to the stricter standard of what it means to be a moral emotion, then it seems unlikely that any emotion does. My analysis of guilt will aim to prove that it will remain standing as a moral emotion in the traditional sense. In the next section, I will show that, at least when it comes to the cases of merely causal harm to others, then those do not pose a particularly strong argument against the idea that guilt is a moral emotion in a traditional sense.

### **3.2. In defence of a moralised view of harm**

In this section, I will argue that there is indeed more to the cases of merely causal harm than harm simpliciter. By constructing several imagined scenarios where we would be dealing with true harm simpliciter (cases that no one really discusses), I hope to make it evident that all the cases that the interpersonal approach appeals to are actually cases of moralised harm. I will begin by highlighting the fact that the interpersonal account of guilt makes no conceptual space at all for the possibility that one may feel sincerely sorry for the harm that they have caused to others, without feeling any guilt. At least intuitively, this seems like a viable

option, and to leave no room for it without any further argument is a major oversight. Next, I will move on to analyse what it is that makes the cases of merely feeling sorry for the harm that one has caused different from the cases of causing harm that involve guilt—cases that the interpersonal account usually refers to. As I will argue, even though cases of causal harm do not involve a sense of moral wrongdoing, they are nevertheless linked to moral badness in a more general sense, and, in fact, moral badness seems to enter those cases in various different layers. As I will show, it seems plausible that one takes a moralised view of (the relevant type of) harm and considers the avoidance of such harm as a basis of their deontic duties as well. Secondly, the harm is perceived as undeserved. And thirdly, the basis of desert is likely to be moral goodness. So, even though on the face of it cases of merely causal harm to others may seem like non-moral matters, a closer inspection reveals they may even be considered as triply moral emotions.

There is an important disclaimer that I want to make though—about what, if anything, can be achieved by the discussion that follows. Since the idea that we might be dealing with moralised harm is a possibility that the proponents of the interpersonal view have not really considered, and most of the empirical research on guilt comes from the proponents of the interpersonal view, there are currently no conclusive studies that would have teased out whether people feel guilt about harm simpliciter or a moralised perception of harm. As such, I cannot support my case with real-life examples but will have to rely on imagined scenarios and hope that the reader will share my intuitions about those cases. Such strategy is sometimes criticised as mere armchair philosophy that should not be taken overly seriously. While I agree that appeal to our intuition about certain cases cannot conclusively settle the matter, I would like to highlight that my aim in this dissertation is not to offer conclusive proof but to provide a plausible alternative explanation that could, in principle, be put to an empirical test. Also, it should not be forgotten that, since I will nevertheless be pointing at the fact that there is an alternative way to explain the cases that have inspired the interpersonal view, the interpersonal view itself will stand on no firmer ground than my own proposal. After all, since they have not considered the possibility that we might be dealing with moralised harm, they too have not conclusively shown that guilt is about harm simpliciter. So, the account that I will offer can hardly be worse than the interpersonal account.

I will argue that the cases of merely causal harm to others that are used to argue in favour of the interpersonal view are not really sufficient to make the case that guilt is about harm simpliciter. As such, I will offer an alternative explanation that puts the emphasis on the fact that the relevant harm must be considered morally bad.

### 3.2.1. Making room for feeling sorry without guilt

One of the biggest failures of the interpersonal account is that it leaves no conceptual space whatsoever for the possibility that one might feel truly sorry for the harm that they have caused, and sincerely empathise with the other person to whom they have caused harm, and yet feel no guilt. If guilt is to be construed in terms of simple harm to others, then, by definition, all instances where one perceives oneself as having caused harm and empathised with the other person must be counted as instances of guilt. The idea that harm, or empathetic identification with another person's suffering, alone is not sufficient to elicit guilt is not itself new. It is partly also for this reason that the proponents of the traditional account have been dissatisfied with the idea that guilt is just a matter of harm to others. The importance of "just feeling sorry for the person one has harmed, or even sorrowful or distressed on his behalf" is also highlighted by Patricia Greenspan (1995, p. 128). And John Deigh is even more thorough when making this case by trying to construct a more concrete example. As I will later explain, I have many reservations about the example that he uses to make the case, but it will nevertheless be a useful starting point, and I am sympathetic to the general intuition that one can empathetically identify with the suffering they have caused and yet feel no guilt. The example that Deigh uses in order to make his case is the following:

Collecting your baggage after a flight, you mistake a stranger's suitcase for your own and start towards the exit with the wrong suitcase in hand. It's an honest mistake, as we say, about which you may, nonetheless, once it comes to light, feel bad for having made and for whatever distress you may have thereby caused the owner of the suitcase to suffer. And if you think that it was an honest mistake, then, though you feel bad about having made it and wish you had not, you need not feel any guilt, nor would there be anything untoward if you didn't. (Deigh, 2019, pp. 54–55)

He describes a case where one accidentally picks up another person's suitcase and starts to head off with that. Soon they realise their mistake and empathetically identify with the distress that they have caused to the real owner of the suitcase. Yet Deigh does not think that the person would necessarily feel guilt. He takes this as a point in favour of the traditional account—that unless one perceives oneself as being morally at fault, one cannot feel guilt. And suggests that the emotion felt in this case is better described in terms of regret rather than guilt. The person feels bad about the deed, wishing that it would not have happened, and empathetically identifies with the distress that he has caused to the suitcase owner, but there is no discomfort with oneself.

I find this a complicated case. First, I am not sure that everyone shares the intuition that one would not feel guilt on this occasion. Taking another person's suitcase may well be considered morally problematic. Even if one does not do so intentionally, one may nevertheless be guilty of negligence. It could be argued that everyone is morally responsible for making their suitcases identifiable in the

first place and for later making sure that the one they pick up from the belt does indeed belong to them. Secondly, the guilt in this case may not be that visible because of its low intensity. The wrong in question is a relatively minor wrong because it has not caused much damage yet. If one would go home with the suitcase and only then discover that it is not theirs (and for the argument's sake, let us assume that one's emotions about taking another person's suitcase are not overshadowed by one's concern for one's own suitcase), then one would have to blame himself for a greater wrong. Third, the guilt may also be neutralised by the fact that the possibility of compensation is close at hand. The person discovers the mistake already on their way to the exit, and it is easy to go back to the conveyor belt and rectify the mistake. By the time they reach the owner, they have already repaired the damage to a great extent. For all those reasons, I find this example somewhat problematic, but I think it is definitely conceivable that the mere fact that one empathically identifies with the distress that they had caused to the suitcase owner does not necessarily have to amount to guilt if the person does not construe the situation in any sense morally problematic. This is important because it is not like one would have to be totally indifferent in order to feel no guilt, the lack of guilt feelings may go hand in hand with sincerely feeling sorry. And this is a serious problem for the interpersonal account that interprets guilt in terms of mere harm to others.

In order to strengthen the case, let us try to come up with some alternative cases where one is the cause of another person's suffering, empathises with another person's situation, yet feels no guilt. Imagine that Victoria rented her apartment to Amanda because she was to spend an exchange year abroad. Now, Victoria comes back and sees that Amanda has become really attached to the place and is very sad about moving out. Victoria, who was sad to move out of her apartment abroad as well, fully identifies with Amanda's situation. She feels sorry for Amanda. Yet she feels no guilt because it was their agreement since the beginning that Angela can only live there until Victoria returns. There is a clear sense of causal responsibility—it is because of Victoria that Amanda has to move out, and there is empathetic identification with the suffering that Victoria causes to Amanda, yet there need not be any guilt feelings.

Also, assuming that there are no complicating background factors, a professor need not feel guilt for giving a student a bad grade, even though they may feel sorry for them because, for example, it ruins their chances of graduating *cum laude*. There may, of course, be cases where the professor does feel guilty, perhaps because of not encouraging the student enough or because of some other reason. But it is, for sure, conceivable that they may simply feel sorry without any guilt.

Let us also imagine a case where we have the sort of Kantian who takes it as their highest duty to be honest and respect other people's autonomy, and does not really think that morality is a matter of promoting happiness. They go out with their friend and admit that they find their friend's dress somewhat cheap-looking. The friend is hurt by such a comment. I am tempted to suspect that even if the Kantian feels sorry for the fact that they have inadvertently made the friend sad, they need not feel any guilt.

As such, I take it that we already have a sufficient number of scenarios to make the case that it is highly plausible that causal responsibility and empathetic identification with the suffering that one has caused is not sufficient for guilt, and may also amount to simply feeling sorry. In what follows, I will try to explicate which factors those are that are missing in the cases discussed above but present in those cases of causal harm to others that we do tend to describe in terms of guilt.

### 3.2.2. Harm and the bad

Before I move on to trying to capture the differences though, let me add a brief note on the notion of harm. Namely, it may, on the face of it, seem better to substitute the talk about immorality with the talk about harm because harm may seem like a more simple and naturalised term. What people consider moral or immoral varies greatly, whereas reference to harm seems clearer. This may be an illusion, though. When we look at the philosophical theorisation on harm, we find so many disagreements that one may begin to doubt whether any single definition of harm could be fit to do justice to the vast array of cases where we are used to talking about harm. Most notably, Ben Bradley, a philosopher who had previously (2009) also attempted to defend a unified account of harm, concludes his work on harm with utter scepticism: “My diagnosis is that the notion of harm is a Frankensteinian jumble. Thus, it is unsuitable for use in serious moral theorising. It should be replaced by other, more well-behaved concepts, such as the axiological concepts of intrinsic and extrinsic badness.” (Bradley, 2012, pp. 54–55)

In fact, most accounts of harm are more concerned with what it means to *cause* harm rather than what constitutes a *harm*. This is what Bradley means when he speaks of extrinsic badness—something being bad because of its effects—and suggests that this is what an account of harm should focus on because what counts as harm in an intrinsic sense is too controversial of a matter (Bradley, 2012, p. 392). So, most accounts of harm are concerned with what it means to harm in the sense of causing harm. For example, one of the most popular views, the counterfactual comparative account of harm, states that an event is harmful to the subject if and only if the subject would be worse off than they would have been, had the event not happened.<sup>61</sup> Most part of this definition is basically just a re-statement of the counterfactual account of causation<sup>62</sup> but could hardly be considered to give us any substantial idea of what counts as a harm (in an intrinsic

---

<sup>61</sup> For example, this view has been proposed in the following works (Bradley, 2009, p. 65; Feinberg, 1984, pp. 33–34; Kagan, 1998, p. 84; Norcross, 2005, p. 150; Parfit, 1984, p. 69).

<sup>62</sup> Unfortunately, the fact that under the guise of harming, authors have simply been talking about causing harm, seems to have gone largely unnoticed. To an extent, this has resulted in some harm theorists trying to reinvent a bicycle that could already be borrowed from the work done on causation. For example, with a background on the theory of causation, it should be bizarre to find that, in the work on harm, authors contrast the counterfactual comparative account of harm with the causal account of harm because, in the literature on causation, the counterfactual account is just one of the most popular accounts of causation.

sense)—it is simply said that the subject will be worse off, which basically just takes us back to the axiological concept of the bad (the state of affairs is worse than before), and what counts as bad is, of course, a matter of great controversy. If, ultimately, the notion of harm refers back to the bad, then it can hardly be said to offer us a clearer alternative than the more fundamental terms from moral discourse because badness is precisely one of those terms.

So, ultimately, what counts as harm in an intrinsic sense is reducible to what is intrinsically bad for the subject. As such, all the disagreements that we have about what counts as bad will transfer into our understanding of what counts as harm. A proponent of hedonistic utilitarianism may not consider a violation of autonomy a harm at all if it makes no one unhappy. For example, if they lie to the birthday child in order to set up a surprise party, they may quite literally see no harm in that. Whereas a Kantian may see it as a great harm because the birthday child's autonomy was violated. It seems that what the proponents of the interpersonal account of guilt usually have in mind is harm construed in hedonic terms—they talk about empathetic pain about another person's suffering (understood broadly, either as physical pain or mental anguish). But if this is the case, then it might be doubted whether this is the kind of harm that elicits the guilt feelings of those who prefer a different moral theory. Some more recent authors are also more explicit in including autonomy violations in the sphere of the interpersonal account of guilt (Mancini & Gangemi, 2021, p. 10).<sup>63</sup>

In addition to the idea that what counts as harm must count as bad, I find it plausible to argue that guilt, in particular, is not just concerned with the bad but more specifically with the morally bad. Let us return to the case of the Kantian who may feel sorry about hurting their friend's feelings, yet not feel any guilt about it. Admittedly, the case is hypothetical. Unfortunately, I have not encountered any relevant studies about the moral emotions of Kantians, but I find it highly plausible that the kind of Kantian who does not think of morality in terms of promoting happiness, need not report any guilt for the mere fact that they have caused emotional distress. For this reason, I find it plausible to suggest that people who report guilt feelings about merely causal harm consider hedonic well-being as an important value, which is also the basis of their deontic duties. A Kantian, on the other hand, may agree that suffering is bad and constitutes a harm, yet they may not take a moralised view of this harm because they do not construe their deontic duties in terms of such harm. Coming back to the earlier case concerning the hurtful comment of the Kantian, the Kantian may indeed find it unfortunate that they have hurt their friend's feelings, and they may be sorry for the mishap, yet the feeling need not amount to guilt because they do not take a moralised view of happiness. If my line of thought is correct, then it is important to note that the person who feels guilt for merely causal harm to others is not really concerned with harm simpliciter to begin with but must take a moralised view of harm. The

---

<sup>63</sup> Strictly speaking, they associate it with 'altruistic guilt', but their notion of altruistic guilt is just a correlate of the type of guilt that the interpersonal account of guilt is supposed to capture.

harm must be seen as morally bad in the sense that what the person considers as morally wrong, too, must be built on that value.

### 3.2.3. Undeserved harm

So, one important factor that may account for the difference between guilt and merely feeling sorry for causing pain, concerns whether one takes a moralised view of pain or not. Secondly, I would like to draw attention to the fact that all the cases that are usually listed as cases of merely causal harm seem to involve an element of injustice or unfairness (I will be using those words synonymously). Even if the subject does not view themselves as violating a moral duty, the harm that is brought upon the victim is nevertheless seen as unfair. For example, if one kills another person in a car accident, then such great harm is nevertheless seen as unfair towards the victim who had done nothing to deserve such fate. And in reverse, without any sense of unfairness, guilt seems less likely. There was nothing unfair about the fact that Amanda had to move out of Victoria's apartment—it was what they had agreed upon. There was also nothing unfair about the student receiving a bad grade—it was a fair assessment of their work. It is because of a lack of unfairness that people feel sorry at best but no guilt about the harm that they have caused.

Reference to matters of desert is not completely unheard of in the context of the interpersonal theory of guilt. There are some later developments of the interpersonal theory that hint at the possibility of unfairness. When Francesco Mancini and Angela Gangemi (2021) construe their notion of altruistic guilt, which is supposed to have the same extension as guilt understood in terms of the interpersonal account, they say altruistic guilt arises from having harmed “an innocent victim” (p. 1). Also, as we will see later on, some of the research provided by Baumeister may point towards the possibility as well because he notes that people do not feel guilt about harming others in cases where they have low esteem towards the other person (Baumeister & Wotman, 1994, p. 132). But those are just vague hints that are usually not further elaborated on.

Another way how a reference to desert might enter into the interpersonal approach to guilt is via its strong link to empathy. Unfortunately, our contemporary talk about empathy as a general term for putting oneself in other people's shoes and understanding their thoughts and emotions, has become somewhat disconnected from matters of desert. Yet when we look back at the history of philosophy, we can see that already Aristotle associated our fellow feelings for other people with matters of desert. In particular, when Aristotle speaks of compassion,<sup>64</sup> which is exclusively understood in terms of identifying with other people's suffering (rather than their positive emotions), he defines it as “a certain

---

<sup>64</sup> Some translations of Aristotle refer to ‘pity’, but I am following authors such as Nussbaum (1996, p. 31), Blum (1987, p. 233), and Kristjánsson (2006, p. 89) in choosing to speak in terms of ‘compassion’ rather than ‘pity’ because, nowadays, the latter tends to be associated with an element of superiority.

pain at an apparently destructive or painful evil happening to one who does not deserve it” (Aristotle, 1991b, p. 152 [1385b], see also Kristjánsson, 2006, p. 90). The landscape of empathy-related terms is, of course, complicated (see, for example, Coplan, 2011), and it is by no means my aim to try to disentangle its nuances, but what is important in the context of guilt is that our emotional identification with the harm that is suffered by other people has indeed been traditionally viewed in connection to undeserved harm, and this ancient wisdom seems to be well aligned with the cases that are usually discussed in the context of guilt as an interpersonal phenomenon.

But given that nowadays, the topic of justice remains largely unthematized in the context of guilt, let us start by clarifying what kind of injustice we are dealing with when we talk about desert. In this broad overview, I will mainly rely on the works of Joel Feinberg (1970, 1974) and Kristjan Kristjánsson (2006). Feinberg is an important classic source—much of the contemporary talk about desert was reinvigorated by his writings. Kristjánsson, even though he has modestly admitted that he does not “claim any originality” with his general characterisation of desert (2006, p. 48), has nevertheless been acknowledged for providing us with an especially “well-articulated” account of desert, which offers “the most forceful arguments in support of the received view of desert” (Kinghorn, 2021, p. 155). One of the reasons why Kristjánsson’s account is so useful for the current purposes is also the fact that, since he is interested in conceptualising desert in the context of emotions and virtue, he is interested in folk conceptions of justice that give rise to our emotions. And in our everyday thinking, desert plays an important role and can also be encountered in the more cosmic rather than merely deontic form. This interest has not always been shared. As noted by several authors like Alasdair MacIntyre and David Miller, there is a lacuna between our folk conceptions of justice and those that can be found in political philosophy—while desert plays an important role in the former, political philosophers tend to be neglectful of the matter (MacIntyre, 1981, Chapter 17; D. Miller, 1999, p. 131).

So, what is desert? Since desert is a matter of noncomparative justice, let us first see what justice is and how non-comparative justice compares with comparative justice and what kind of forms non-comparative justice can take. Justice, in general, is a matter of people getting their due. Comparative justice determines what is due to a person by reference to other people, while noncomparative justice determines it independently of other people (Feinberg, 1974). Simply put, comparative justice is a matter of consistency: it “requires that relevantly similar cases be treated similarly and relevantly dissimilar cases be treated dissimilarly in direct proportion to the relevant differences between them” (Feinberg, 1974, p. 310). For example, if all children have been equally good, then we cannot just arbitrarily leave one child without a candy while all the rest get one. This would be arbitrary discrimination. But when we are talking about non-comparative justice, what is due to a person is not dependent on how much everyone else gets. If the child has completed the test to a maximum score, then she should get the

highest grade—regardless of what grades are given to other children.<sup>65</sup> If one has a right to life, then the fact that it would be non-comparatively unfair to violate this right does not depend on whether other people’s right to life is violated or not. For our purposes, it is important to differentiate two types of claims that non-comparative justice may involve: entitlements such as rights, which imply that other people have an obligation to respect your rights, and desert claims about what is due to people based on their merits, most importantly, based on their moral goodness.

In the cases of guilt without perceived wrongdoing that are the focus of this dissertation, the victim’s rights are not usually violated because, if they had, then the subject would already be guilty of (at least *pro tanto*) wrongdoing. If it is an accident that someone is killed in a car crash, then it is not the case that one’s right to life has been violated. At least according to a classic understanding, one’s right to life can be violated only if someone takes their life by morally wrongful actions, not by accidents, natural catastrophes or diseases. But, in some rare occasions, guilt without wrongdoing can be attached to the violations of rights as well. This is the case when one is further down the causal chain, so that they are not the direct bringer of harm but the harm is eventually administered by the morally wrongful actions of someone else. Remember the story of the Jewish man who had sat beside Adolf Hitler at the school entrance exam. The fact that he passed the exam but Hitler did not, seemed to have raised feelings of resentment in Hitler, which might have played a role in his later action. This is clearly a case where the victim’s rights were violated because of the morally gruesome actions of the Nazis. Additionally, guilt may also attach to the violations of rights in case of indirect guilt—for example, when someone tries to avenge or manipulate a person by means of hurting their loved ones. In both cases, the person feels guilt for their role in the causal chain that led up to a non-comparatively unjust state of affairs that consisted of other people’s rights being violated.

Most of the time, however, in cases where one is a more direct cause of other people’s suffering, we are dealing with desert claims rather than right claims. So, in the case where someone accidentally kills another person in a car accident, the harm is better described as being perceived as undeserved. Since we are not interested in deontic violations here, the account of desert that is relevant for our purposes is the one that has been called ‘telic’ (Olsaretti, 2003, p. 8) or ‘cosmic’ (D. Miller, 1976). Such an account focuses on states of affairs that take place out there in the world, regardless of how they have come about—be it by the hand of

---

<sup>65</sup> Admittedly, some authors like Kristjánsson, are against the possibility of non-comparative *desert* in particular. He argues that all desert is ultimately comparative because it depends of “‘holistic’ comparisons with the general distribution of the relevant benefits across the society” (Kristjánsson, 2006, p. 62). The idea is that it cannot be a standard for an average person to deserve a fancy mansion if there are not enough resources to grant them such mansions. So, “although it does not mean that they [desert claims] must all involve a comparison with the speaker’s own lot”, they are nevertheless comparative in a wider sense (p. 62). I will be content to use the word ‘noncomparative’ in the narrow sense of not involving a comparison with the speaker’s own lot.

some moral agent (qua moral agent) or not.<sup>66</sup> So, even if the victim was not strictly speaking wronged by someone's culpable actions, the harm that they suffer can nevertheless be viewed as undeserved because desert only concerns whether the person receives the kind of fate that corresponds to their merit. More specifically, desert claims can be analysed as taking the form "S deserves X in

---

<sup>66</sup> At this point, when we shift from the deontic to the cosmic, some may perhaps become suspicious of desert. For some, the phrase 'cosmic justice' may perhaps associate with karmic justice, which refers to the idea that everyone will eventually get their due. This is not the sense in which cosmic justice is discussed in this dissertation. The focus is only on the idea that desert concerns states of affairs that need not be brought about by a morally culpable agent. On a slightly more serious note, it might perhaps be questioned whether what is called cosmic justice is actually just the good old deontic justice in a disguised form. If one thinks about cosmic injustice as the doing of some kind of deity or personification of fate, then it may indeed be possible to think of such injustice as a matter of this deity acting in immoral ways. (Notice that the truth of such claims is beside the point here because we are only concerned with how people perceive matters of morality, as only this will ultimately translate into their emotional experience.) I am not sure how common it is to think in such a way. Even when we talk about people who are religious, there seem to be several beliefs that could block them from attributing outright immorality to the deity they believe in. To begin with, attributing morally wrongful actions to a deity does not seem like the most straightforward route for the proponents of the divine command theory, according to which morality is itself determined by the deity who sets the rules. If what is right is that which the deity loves, and what is wrong is that which the deity hates, then, assuming the deity does not act in contradictory ways, it is almost by definition that whatever the deity does is right and good. For this reason, I assume that those who attribute cosmic injustice to the doings of some deity are more likely to conceive of morality as autonomous from the deity's will. Another obstacle in the way of attributing cosmic injustice to the wrongful acts of a deity is the belief in an omnibenevolent, omnipotent and omniscient god—for it does not seem logically consistent that an omnibenevolent perfect being would do something that is outright morally wrongful and blameworthy. Yet it is easier to see how the gods of the Ancient Greeks and Romans could immorally bring about natural catastrophes and various other misfortunes—because their gods were more humanlike in their ways. And it is also fairly relatable how even people with a more secular outlook might sometimes anthropomorphise fate in a somewhat quasi-religious manner. But even if instances of what we call cosmic justice are sometimes experienced as reducible to deontic justice, I find it highly unlikely that this should always be the case—that all talk about cosmic justice must ultimately collapse into people either consciously or unconsciously construing the situation as a morally wrongful doing of some divine entity. This is because the core element of cosmic justice is by no means tied to the deontic requirement—people can receive or fail to receive their due regardless of whether they are dealt by other moral agents or whether things just happen to them. And for this reason, we can and often do talk about cosmic desert without any hidden reference to deities or quasi-religious entities like fate. Also, it would ultimately just turn into a dogmatic (and unfalsifiable) insistence to suggest that, even in the absence of any evidence that a particular person is indeed thinking in religious or quasi-religious terms, there must nevertheless be some unconscious idea of an agent who brings the injustice about. As I have already explained in section 2.3, I do not think that we can engage in any serious philosophical theorisation if we operate with such dogmatic insistences about people's psychology. Thus, I do not see a reason to outright deny that we can coherently talk about a cosmic concept of justice at all, and that all justice must be reducible to deontic justice.

virtue of  $F$ ” (Feinberg, 1970, p. 61), where  $S$  is a person,  $X$  is a positive or negative state of affairs,<sup>67</sup> and  $F$  are the facts about  $S$  that ground the desert claim, often also called ‘desert bases’. For example, we might say that teachers deserve a high salary by virtue of being dedicated to an important cause or that Matt deserves to find love by virtue of being such a nice and kind-hearted guy. In fact, one of the most common desert bases is moral virtue, but I will leave the discussion about that for the next subsection.

Overall, the idea that guilt for merely causal harm involves a sense that the victim does not deserve to suffer the relevant harm is a highly plausible one because desert seems to be the crucial factor that can prevent people from feeling guilt. It is common among psychologists to underscore that feelings of guilt depend on the esteem that one has for the victim, and one way to immunise oneself from guilt feelings is to derogate the victim (Baumeister et al., 1995a, p. 259ff; Bernard et al., 1965; Lazarus, 1984; Lerner & Matthews, 1967; Miceli & Castelfranchi, 1998, p. 306). In fact, the cases of unrequited love that Baumeister and colleagues refer to as important instances of guilt for merely causal harm follow the same pattern. The rejectors reported no guilt feelings in cases where the would-be lover was seen as ““a total turn-off loser”, unpopular, ugly, and with an unappealing personality”, as “an unethical, malicious, repulsive individual who had deliberately hurt others” or as being “egocentric, self-centred” and having “the personality of an ice-cube” (Baumeister & Wotman, 1994, p. 132). Even though it may be doubted whether ugliness is an appropriate desert base, all the rest of the characteristics that are listed refer to matters of character that are considered rather common features that make a person less deserving. This seems to provide rather strong evidence that the rejectors are not just concerned with how much they have hurt the would-be lover but also with whether the would-be lover is deserving of the gift of love. This is also the case in my previous example of a teacher handing out a bad grade—even though they might be sorry for the student, they need not feel any guilt if they perceive themselves as having every right to give this grade and if they also think that the unsatisfying answers in the test were not just due to some unfortunate accident (e.g. the student being emotionally disturbed by a sudden personal tragedy) but truly did reflect the student’s ability. Similarly, Victoria did not feel guilt for asking Amanda to move out because there were no rights violations involved, nor did she think that there was any reasonable sense in which Amanda should have deserved her apartment in particular. So, given that viewing the harm as deserved is going to prevent guilt feelings, it is only natural to assume that in cases where one does feel guilt, the subject must see the harm as undeserved. This is also well in line with the intuition that we might feel a particularly strong sense of guilt for harming a child because children are considered especially innocent, as they have not had much

---

<sup>67</sup> I am here relying on Kristján Kristjánsson’s modification. Originally, Feinberg mentions “a mode of treatment” but since I’m dealing with a telic concept of desert, it is more appropriate to talk about “a negative or a positive outcome or state of affairs”, as Kristjánsson does (2006, p. 51).

chance to commit serious moral ills that would make them undeserving of good things in life. And it is also aligned with the intuition that guilt is stronger when we have harmed an especially good and kind-hearted person who in no way deserves to suffer the harm.

Overall, there seems to be a strong case in favour of the idea that cases of merely causal guilt are interpreted as being noncomparatively unfair to the victim, where this unfairness usually comes in the form of undeservingness. This makes guilt a partially desert-based emotion, and as such, I would like to finish the section about guilt's relationship with desert by positioning it on the wider landscape of desert-based emotions, which has been mapped out by Kristjánsson. According to him, the basic idea that people should reap as they have sown captures the following constellation of ideas (which I will pair with the corresponding virtuous emotions, as mapped in (Kristjánsson, 2006, p. 100): it is bad if bad things happen to good people (compassion), it is good if good things happen to good people (gratulation), it is bad if good things happen to bad people (indignation), and it is good if bad things happen to bad people (satisfied indignation). Kristjánsson analyses all those responses and also their less virtuous counterparts (like feeling *schadenfreude* rather than compassion in response to good people suffering bad fortune) but since guilt is not exclusively related to matters of desert (traditionally, it is viewed in relation to wrongful action that violates rights), it does not receive a place in Kristjánsson's study of desert-based emotions and is also generally not discussed in relation to matters of desert. But even if guilt is not exclusively related to desert, it is nevertheless partly related because, in cases of causal harm, guilt feelings depend on perceiving the harm as undeserved.

#### **3.2.4. In defence of desert as a moral matter**

So far, I hope to have established a strong connection between guilt for merely causal harm and desert as a form of noncomparative justice. Yet, I take it that some may doubt whether emphasising the link with desert is sufficient to grant guilt the status of a moral emotion. This is because desert is sometimes just seen as a matter of personal worthiness, which has little to do with morality. So, in this section, I am going to strengthen the case in favour of counting desert-based forms of guilt as moral emotions. First, I will respond to the more general critique against desert as a moral concept that has been advanced by authors such as David Miller and Aaron Ben-Ze'ev. This will help to establish that desert, in general, can rightfully be counted as a moral matter. Secondly, I will further strengthen the case for desert-based guilt being a moral emotion by arguing that, at least in cases of guilt for merely causal harm, there is an additional layer of morality introduced by the fact that the relevant desert base must be a moral one.

The problem with desert being a moral matter is closely related to the fact that it represents a cosmic concept of justice, and it might be questioned how much importance we should attach to judgments of cosmic justice. Namely, some authors have been insistent that once we are not talking about rights that would be grounded in the obligations of other people, but shift the focus to desert, we

are not really dealing with a moral matter anymore. In fact, it seems that one of the reasons why David Miller categorises matters of desert as reflecting a ‘cosmic’ sense of justice is precisely that he deems such sense of justice distant and morally irrelevant—it “carries no implications for human actions” (D. Miller, 1976, p. 115). We can also see the same idea reflected in Aaron Ben-Ze’ev’s critique of envy as a moral emotion (Ben-Ze’ev, 2002, pp. 151–152). Identifying the core evaluative concern of envy with the negative evaluation of one’s undeserved inferiority (the subject does not have something that they would deserve), he claims that the reference to desert does not grant envy the status of a moral emotion. Why? Because desert just concerns matters of personal worthiness rather than true moral entitlements. The fact that we do not have something that we deserve is often due to natural differences or other impersonal causes, not because of the morally blameworthy actions of other people, and neither is it necessarily linked with anyone’s obligations to grant us what we deserve. Both Miller and Ben-Ze’ev express the idea that justice is morally relevant only in the deontic rather than the cosmic sense. If this is true, then this would undermine my attempt to show that the relation with desert can explain why guilt remains a moral emotion even in cases where one is merely causally rather than morally responsible for the harm. Because, currently, my argument lies precisely in the idea that guilt remains a moral emotion because the subject nevertheless perceives the harm that they have caused as undeserved by the victim.

But what should we make of the idea that desert is not really a moral matter? Miller and Ben-Ze’ev do not give us any further argument besides referring to the lack of necessary connections with moral obligation. As such, they seem to rely on the assumption that morality, in general, is a deontic matter, which only concerns the rightness and wrongness of our actions. Yet I see no reason why we should adopt such a narrow understanding of morality. In particular, there are two points of criticism that I would like to draw attention to. Those two points link to the fact that there are, in principle, two ways in which a proponent of exclusively deontic morality can view matters of justice as related to moral obligation—through backward- and forward-looking responsibility. The deontic sense of justice links to backward-looking responsibility to the extent that, for something to count as unjust, it must be brought about by other people’s violation of moral obligations—retrospectively, we can say that someone was morally culpable for the unjust state of affairs. Yet the deontic sense of justice can also link to forward-looking responsibility to the extent that, for a state of affairs to count as unjust, it must be considered as giving rise to someone’s obligation to do something about the matter. My aim is to show that even though matters of desert may lack a tight connection with both backward- and forward-looking responsibility for our actions, they are nevertheless linked to other backward- and forward-looking moral elements that warrant labelling desert a moral matter.

Let us first look at the fact that calling something—for example, a family of good people losing their house due to some unforeseen natural catastrophe—undeserved does not necessarily imply that someone bore backward-looking moral responsibility for bringing this state of affairs about. Even if there is no one

morally responsible for this unfortunate state of affairs, that is, no one who *ought not to* have brought the catastrophe about, the matter is not altogether disconnected from the realm of ought—there is nevertheless a sense in which this *ought not to* have happened. Even though we are not dealing with our sense of what someone ought to do, we are still dealing with the sense of how things ought to be.<sup>68</sup> And the reason why good people losing their home is not how things ought to be is that it is unfair—what people get is not in accordance with what they deserve. In that sense, such a world is nevertheless morally out of order.<sup>69</sup>

But how about calling desert normative in a more general sense rather than moral? At least this is what Aaron Ben-Ze'ev seems to suggest in his critique of envy as a moral emotion.

Envy is surely not a moral emotion, since its core evaluative concern—namely, the negative evaluation of our underserved inferiority—is not moral. This concern does not express a moral concern regarding how people should be treated, but merely a personal concern, regarding what our fortune should be. Although this is not a moral concern, it is a normative one. All emotions involve a normative claim, which is expressed in their evaluative component, but this claim does not have to be moral. Indeed, envy includes a normative claim—it argues something about how situations of fortunes ought to be. However, this is not a moral claim since it does not refer to the way we should treat other people. (Ben-Ze'ev 2002b, 151)

Ben-Ze'ev is careful to note that even though concern for one's undeserved inferiority does not render envy a moral emotion, there is nevertheless a sense in which it reflects a normative concern in a more general sense. But the kind of normativity that he has in mind is characteristic of all emotions because, as he himself points out, emotion types in general can be identified in terms of different core evaluative concerns and thus involve an evaluative component.<sup>70</sup> But this is not enough. The way in which fear, for example, is normative is very different from the way how desert-based emotions are normative. In the case of fear, the subject is concerned about being in danger. Fear is normative in the sense that it involves a negative evaluation of the situation at hand as dangerous—being in danger is considered to be bad, something undesirable. Yet there is more to desert-based emotions<sup>71</sup> than simply perceiving some state of affairs as bad or undesirable. If one is truly concerned about one's undeserved inferiority, which

---

<sup>68</sup> This idea has also been pointed out in Marguerite La Caze's response to Ben-Ze'ev's critique of her idea that envy can be a moral emotion (La Caze, 2002, p. 157).

<sup>69</sup> I will have more to say about the kind of evaluative framework that involves a sense of how things ought to be in chapter 5.

<sup>70</sup> He is using the term core evaluative concern more or less synonymously with Richard Lazarus's notion of 'core relational theme'.

<sup>71</sup> I will have more to say about desert-based emotions in the sense of desert-based emotion types later on in this subchapter. Here I am using the word in the more general sense, applying to any desert-based emotion token (regardless of whether the relevant emotion type in general is necessarily desert-based, meaning regardless of whether all the emotion tokens of the relevant emotion type involve a reference to desert).

Ben-Ze'ev links to envy, they do not just feel that it is bad that they are inferior, they also have a sense that they ought not to suffer such a state—it is unfair, they deserve better. This is something stronger than simply perceiving the situation as bad in a more general sense, like in the case of evaluating the situation as dangerous in case of fear. Of course, one might try to rephrase the fear example and suggest that, in case of fear, too, one might characterise the dangerous situation as something that should not have happened. Confrontation with a dangerous wild animal might be characterised in terms of the subject feeling “Oh no, this should not be happening right now! Must find a way out of here!”. But any such attempt to translate the emotional experience of fear into the language of ought is, at best, bound to capture a prudential sense of ought, a practical ought that mostly just concerns our own well-being. The kind of ought involved in desert claims is of a different kind. It does not just concern the subject's well-being but reflects a more disinterested concern for the world being ordered in a just way.

This special sense of ought that our talk about desert implies has also been aptly captured by Joel Feinberg:

The noncomparative conception of injustice is sometimes implied, I think, in *our talk of states of affairs as “not right.”* It is part of the conventional wisdom in Anglo-American analytic ethics that the distinction between “right” and “good” in ordinary language consists partly in the fact that “right” applies to actions only, whereas “good” is used more generally to appraise not only actions and things but any state of affairs.<sup>8</sup> On the contrary, we do sometimes speak of *states of affairs as right or wrong*, and when we do, we do not intend to say merely that they are good or bad in different but equivalent language. We say, “It is not right that such and such should be the case,” and this is *a stronger and sharper complaint than simply stating that it is not a good thing that such and such is the case.* [---] When we determine that a state of affairs as such is not right, or that *the universe of which it is a part is “out of joint”* in something like the Platonic fashion, without reference to the claims of wronged parties, our judgment is noncomparative. (Feinberg 1974, 309; HM emphasis in italics)

Feinberg is indeed pointing at an interesting possibility—he sheds doubt on the very distinction between deontic wrong and evaluative bad (to the extent that the latter is to be understood as merely undesirable) by arguing that there is room to a special sense in which certain states of affairs are not right and ought not to be—a sense that can be explained by appeal to a telic conception of justice.

It is because of desert's connection with this special sense of ought that it makes sense to categorise matters of desert as moral rather than nonmoral matters—their intimate connection with a disinterested ought justifies viewing them as belonging to the same wider category with matters of more traditional deontic morality rather than the nonmoral sphere that is mostly governed by prudential concerns. To take stock, even if Miller and Ben-Ze'ev are right to point out that, compared to deontic obligations, matters of desert have a less important role in morality, this should not be taken as a ground to automatically label them nonmoral. Especially in the sphere of emotions, which all involve some evaluative

concern, it should be clear that emotions that concern us with justice—come it in the form of entitlement or desert—are nevertheless very different from other emotions that are mainly just concerned with our own well-being. And for this reason, it makes sense to call them moral rather than nonmoral.<sup>72</sup>

By now, we have established that calling something deserved or undeserved nevertheless tells us something about the moral orderliness of the world, even if this orderliness is not brought about by the morally culpable actions of other people. As such, it does involve a backward-looking element that belongs to the sphere of morality—it involves evaluating the state of affairs that has come about as undeserved. But what about the connection between desert and forward-looking responsibility? Is desert morally irrelevant in the sense that it lacks a necessary connection with what we ought to do about the situations that are undeserved? First, it is useful to note that the fact desert claims do not always generate an all-things-considered duty to do something about the matter does not necessarily mean that they cannot generate *pro tanto* duties that might be outweighed by other factors (either other types of justice or other moral concerns), or *prima facie* duties that may simply lose their force because there is nothing we can do about the matter. A similar line of thought seems to be expressed by Kristján Kristjánsson when he responds to Miller’s doubts about the moral relevance of desert by discussing the moral relevance of hypotheticals, stressing that “much of moral development and moral reflection takes place in the God’s-eyes-view realm of the hypothetical: ‘It is undeserved that this innocent child is dying of an incurable disease, and if we could, we should be doing something about it.’” (Kristjánsson, 2006, p. 59) Secondly, even if desert claims do not generate a perfect duty to adjust the world according to desert, they could nevertheless be viewed as providing moral reasons to do something in the sphere of the morally praiseworthy. A researcher who is dedicated to finding a cure for some horrible disease can nevertheless be characterised as taking part in a morally

---

<sup>72</sup> All the being said, I still think that Ben-Ze’ev is correct in his final verdict on envy though—that envy as such is not a moral emotion. I simply disagree about why this is the case. In particular, I think it is an overmoralisation of envy to define it in terms of undeservingness, and we would do better to simply view undeservingness as an intensity variable—something that makes us have a stronger emotional reaction. In fact, it is Ben-Ze’ev himself who lists desert as an intensity variable in his book “The Subtlety of Emotion” (Ben-Ze’ev, 2001, p. 146 pp). Furthermore, we should also take into account that a person in the grip of envy may indeed be likely to evoke the idea that his inferiority is undeserved but this need not reflect the fact that he actually cares about the kind of cosmic justice that desert reflects—this is because he may just employ the concept of desert for self-serving purposes, in order to compensate his self-worth and bolster his claim for the relevant goods. So, even though he himself may employ a moral concept when accounting for his envy, this does not mean that his emotion is actually driven by concern for cosmic justice because he may just be self-deceived. As such, it might be appropriate to call envy a pseudo-moral emotion, which may present itself for the subject as driven by noble goals but which in fact is not. And how do know that it is not? (Read: are we not lapsing into the realm of pseudo-scientific dogmatism?) By observing how the subject reacts when he encounters other instances of people receiving undeserved fate. Those reactions may reveal that he does not care that much about desert at all.

laudable endeavour even if he is under no obligation to dedicate his life to such a cause. Thirdly, even if the fact that something is deserved or undeserved does not have any implications on our actions, it may nevertheless be relevant to our emotional reactions. As pointed out by Kristjánsson, these are not only actions that are morally important but emotions as well—“desert claims are very often emotion claims, grounded in a desire for ‘poetic justice’, rather than direct incitements to action” (2006, p. 59). So, even if one’s judgment of desert (or the lack of it) does not always translate into action, it would nevertheless be a sign of virtue and moral maturity if one is disappointed at an unworthy person winning a prize and a good person suffering an undeserved fate. In other words, even though desert claims do not necessarily concern what we ought to do, they do have implications for which emotional dispositions we ought to cultivate.

Thus, though judgments of desert are not strictly deontic, they are still moral matters in a wider sense. Even though they do not involve a backward-looking deontic judgment of moral wrongdoing, they do involve a backward-looking telic judgment that deems the state of affairs as not simply undesirable but not right in the sense of not being in accordance with people’s merit. And even if judgments of desert do not generate a forward-looking deontic duty to, all-in-all, do something about the matter, they can still generate a *prima facie* or *pro tanto* duty, and even if there was no deontic duty that concerns our actions, judgments of desert may still have implications on how we ought to feel.

All the above concerned desert in general, regardless of the desert base. Saying that both deontic desert and cosmic desert are moral matters is not dependent on the desert base being a moral feature because the focus is solely on people getting what is due to them, not on in virtue of which feature exactly it is that something is due to them. This is so when we talk about deontic desert, in which case a person deserves a certain form of treatment by others. Imagine Anna, who is highly intelligent and contributes greatly to the company’s success, but her boss systematically underestimates her ability because he thinks that women are more led by their emotions than reason and never gives her the proper recognition for her achievements. Surely the boss would be a misogynist, and there is no doubt that misogyny is an important moral failure—just like other vices of superiority<sup>73</sup> like racism. The fact that the desert base is Anna’s intelligence, which is a non-moral matter, does not make any difference to the fact that desert itself is a moral matter, and that it would be morally right to recognise Anna’s intelligence and achievements. Joel Feinberg would label the case as one of ‘judgmental injustice’, where the injustice (considered apart from any harmful consequences that it may have for the victim) consists simply in the falsity of the boss’s derogatory judgment (Feinberg, 1974, p. 302). As Feinberg puts it, the moral offensiveness of such discrimination is *sui generis*—it derives from “its absolute groundlessness”, the fact that it is ““offensive to reason”—absurd, arbitrary, disproportionate, or inconsistent” (1974, pp. 318–319). Also, if the desert base does not matter for categorising deontic desert as a moral matter, then it should not matter for

---

<sup>73</sup> I lend the phrase ‘vices of superiority’ from (Bell, 2013, p. 96).

categorising cosmic desert as a moral matter either. Imagine a slightly modified case where Anna does not get any recognition at the annual Christmas party, even though this is the venue where the best employees are acknowledged, and Anna definitely deserves to be recognised for her great achievements. This lack of recognition may be due to a misogynist boss, but, given that we want a case where no deontic error would be involved, let us imagine that it was instead due to an accidental error that occurred when the list of best employees was passed on to the organisers of the event (perhaps the ink ran out just when the printer was about to print Anna's name or something alike). Even if there was no moral error involved, it is nevertheless unfair that Anna does not get recognised at the Christmas party because it is not in accordance with her merit, and thus there is reason to do something about the matter (perhaps try to inquire about why Anna was not mentioned, or to make a speech in honour of Anna's great contributions oneself), or at least be disappointed that justice was not served.

Last but not least, I would also like to draw attention to the fact that at least in the context of guilt for merely causal harm, the vast majority of cases will rest on a moral desert base, and as such, include an even further layer of moral concern. But, first, let me be clear about the function that this idea serves in my overall case in defence of guilt as a moral emotion. I take it that, by now, I should have already made a sufficient case in defence of desert in general being a moral matter, which will already provide sufficient grounds for counting guilt for merely causal harm as a moral matter, given that those cases are about undeserved harm. As such, my main case in defence of guilt as a moral emotion has already been made, and my point about the prevalence of a moral desert base should be taken as merely pointing toward an additional layer of morality that is not necessarily omnipresent but nevertheless there in the vast majority of cases of guilt for merely causal harm, especially the cases like car accidents and heartbreaks, which are some of the central examples of guilt for merely causal responsibility. To the extent that those examples have been presented as nonmoral instances of guilt, it would be an important oversight not to point out this further layer of morality because it goes to show that we should not underestimate the role that morality can play in seemingly simple judgments of harm. In fact, it is quite amazing how something that may first seem like a nonmoral matter may, upon closer inspection, reveal multiple layers of moral concerns.

Before I go on to my own more modest argument that is only meant to apply in the context of cases that involve harm to others, and even there does not have the pretension of covering absolutely every case, I would first like to draw attention to a more ambitious strategy. Namely, some authors suggest that all desert claims rest on moral virtue as the desert base. It is not the contemporary majority view and has been criticised as “a characteristic mistake of philosophers” (D. Miller, 1999, p. 134), but the idea can be found in the works of classic authors like Ross (2002, p. 138) and has been the most elaborately defended by Kristjánsson (2006, pp. 64–79). Even though it would benefit my cause if all desert claims rested on the single basis of moral virtue, and even though overall, I do think that Kristjánsson makes some rather compelling arguments, I am not

convinced that it is never right to talk about nonmoral desert bases. In particular, I am sceptical about Kristjánsson's analysis of desert in the context of winning competitions. In a memorable example, he points out that Michael Schumacher, a once successful Formula 1 driver and a multi-millionaire, did not deserve to win any Grand Prix races because he showed no concern for his grandmother, who lived in dire straits and never received visits from him (2006, p. 65). This is supposed to illustrate the importance of virtue as a desert base. And the idea that Schumacher deserved to win in virtue of his excellent driving skills is put aside as a 'sham' desert judgment<sup>74</sup>—a kind of loose talk about 'desert' in our folk discourse that is actually about entitlement rather than desert and means that according to the rules of the institutionalised practice of racing competitions, Schumacher had the right to be crowned as the winner (just like according to the rules of the school system, students may be entitled to receive a good grade). I think that Kristjánsson may be right in the overall judgment that, as a person who had totally neglected his grandmother, Schumacher did not deserve all the external rewards that came along with winning. But this is compatible with the idea that, as an excellent driver, he did nevertheless deserve to be acknowledged as the winner. Because his wins were not just a row of lucky chances that were merely permitted by the rules,<sup>75</sup> but he did really exhibit the kind of sporting prowess that the relevant competitions were designed to test. In other words, his wins were not only in accordance with the rules of the game, but they were in accordance with the spirit of the game. So, while his driving skills, or his driving skills alone, might not have been sufficient to make him worthy of all the money and the fame, there is still room for saying that he nevertheless deserved to be acknowledged as the best driver.

I am sympathetic to the idea that, in a pre-institutional sense, it is not clear that nonmoral talent as such can make people deserving of external goods—understood as material rewards and good fortune. As also put by David Miller (1999): "I can mark out a 400-meter track and run around it in 45 seconds, but in the absence of an institution of competitive sport there is nothing (except perhaps a kind of bewildered admiration) that I deserve for doing so" (p. 139). Or imagine an artist just painting to themselves and never sharing their art with the world—here, too, it does not seem that talent alone will make them deserve anything in particular. So, to the extent that we are concerned with external goods like material rewards and good fortune, Kristjánsson might be right to say that it is moral virtue rather than non-moral virtue that can make us deserve those things. The idea that desert claims are founded on virtue does indeed have great intuitive appeal. It relies on a culturally universal deep-seated pre-theoretical idea that one should reap as they have sown (Pojman, 2001, pp. 98–99; Kristjánsson, 2006, p. 57), which, as Kristjánsson rightly points out, is something that we are taught

---

<sup>74</sup> Kristjánsson borrows this phrase from David Miller (1999, pp. 137–138).

<sup>75</sup> Like, for example, another case that Kristjánsson discusses (2006, p. 46)—the case of Stever Bradbury, who won the 2002 Winter Olympics in men's 1000 metre short-track speed-skating because all the rest of the five competitors fell on the final bend.

since early childhood—even fairy tales are often built around the idea that the good guys should win, and the bad guys lose. But, in completely denying that nonmoral virtue, too, could serve as a desert base, Kristjánsson seems to overlook the sphere of desert that Feinberg calls judgmental justice, where the injustice does not rely on an unfair distribution of goods but in the mere “falsity of the derogatory judgment” (Feinberg, 1974, p. 302). And I think that this is an important form of justice. The vices of superiority, like misogyny and racism, are in great part founded on judgmental injustice—both the misogynist and the racist underestimate other people not so much for being immoral but first and foremost because they perceive them as inferior by some nonmoral measure like intelligence. For those reasons, I do not want to start from flat-out denying the possibility of nonmoral desert bases.

However, I do think that there is a case to be made for the general prevalence of moral virtue as a desert base. This is also attested by the fact that empirical studies seem to confirm that the predominant desert base is moral virtue (Lupfer & Gingrich, 1999). And the case for the prevalence of a moral desert base seems especially strong when it comes to cases of merely causal harm, like those of car accidents or causing heartbreak. There are three aspects about those cases in particular that make a moral desert base especially likely. First, to the extent that non-moral desert is most closely related to judgmental justice and cases of merely causal harm to others are not instances of judgmental injustice, it seems that they are more likely to be moral rather than non-moral matters. They involve a more substantial harm that usually comes in the form of physical pain (as in case of car accidents) or mental anguish (as in cases of heartbreak). The reason why they do not involve judgmental injustice is probably the fact that it is quite hard to accidentally cause judgmental injustice. It would have to include accidentally causing someone else to think badly about the victim’s non-moral merits, which could perhaps only happen if one naively passes on false information. So, building from the insight that cases of merely causal harm usually involve reference to more substantial goods, and combining it with the assumption that desert that rests on a nonmoral desert base is not a matter of more substantial goods but only other people’s favourable judgment (with regard to the relevance competence), then cases that concern more substantial goods (or the lack of them) do not seem like particularly good candidates to involve a reference to a non-moral desert base.

Secondly, to the extent that someone might still want to insist that nonmoral virtues can make people deserving of more substantial goods, I would want to highlight that those more substantial goods are then likely to be some kind of awards that only have a place within an institutional context. But cases of causal harm to others are not usually the kinds of cases where one would prevent the victim from receiving an award. And nor is it particularly simple to accidentally prevent someone from receiving a reward because institutionalised contexts that function on the basis of non-moral merit are usually quite controlled environments, where it is not that simple to accidentally prevent someone from getting what they deserve. Even the example that I used earlier, the one where Anna did

not get recognised at a company's Christmas party, was a rather unusual case—in order to make the harm accidental, I had to add the assumption that her name got lost when the list of people that were to be praised for their high contributions got lost because the printer ran out of ink just before her name got printed. The person who forwarded the incompletely printed list may indeed have grounds for merely causal guilt, so I am not denying that there can be some rare cases. I am simply arguing that those cases are bound to be rare.

Thirdly, and most importantly, cases of merely causal harm characteristically involve not just the lack of certain goods but outright harm, usually in the form of physical or mental pain. And morality seems like the only sufficiently serious criterion based on which this kind of harm can ever be viewed as deserved. Nobody thinks that people deserve to suffer harm simply because they are not particularly talented in maths or great in sports. This is confirmed when we look back at why it was that people in Baumeister's own study did not feel guilt for breaking another person's heart. They describe the would-be lover as "an unethical, malicious, repulsive individual who had deliberately hurt others", as being "egocentric, self-centred" and having "the personality of an ice-cube", as "a total turn-off loser", unpopular, ugly, and with an unappealing personality", (Baumeister & Wotman, 1994, p. 132). Admittedly, ugliness and being a loser are nonmoral matters, but all the rest of the characteristics that are listed refer to matters of moral character. Furthermore, even though moral virtue as a desert base is already common in cases where the subject did not feel guilt, there is reason to believe that it will be even more common in cases where the subject does feel guilt because people who do feel guilt are generally morally more sensitive than the ones who manage to prevent guilt feelings by derogating their victims by all kinds of criteria like intelligence or skin colour. When one already feels guilt about non-culpably causing the would-be lover's heartbreak, it does not seem plausible that they would think that the victim does not deserve the anguish of heartbreak because they are, for instance, talented in mathematics. A much more plausible explanation is that they are concerned with hurting a good and kind-hearted person.

Of course, we do not always know about the moral virtue of the victim. For example, in cases of car accidents, the victim may be a stranger. Is there any reason to suppose that, in those cases, too, guilt feelings involve the sense that the harm is undeserved because of moral virtue? I believe so. It does not seem plausible that one would view the death as undeserved because the victim was, for instance, a great mathematician or worthy by some other nonmoral merit. If there are people at all who do not deserve to live, then those would be the most gruesome of criminals. So, one does not have to be a great mathematician to deserve to live, they just need to be a sufficiently decent human being. And even if the person who feels guilt does not have full knowledge about the virtue of the victim, this kind of common moral decency is something that we can assume by default. Moral monsters are rare, and in the absence of any reason to think that the victim was exceptionally vicious, it is safe to assume that the victim had done nothing to deserve such a fate.

This concludes my modest case in defence of the idea that cases of merely causal harm have a predominantly moral desert base. The desert base is moral because those cases usually take place in a non-institutional context where a more primordial sense of desert is at play, which simply relies on the idea that it is unfair for the good guys to suffer great misfortunes.

Let us take stock. Even though, at first sight, cases of guilt where one is only causally rather than morally responsible for harm to others may seem like nonmoral instances of harm simpliciter, I hope to have shown that this is far from being the case, and there are multiple layers of morality built into the kind of harm that plays a role in those cases. Let me summarise all those different layers. First, the notion of harm itself is ultimately reducible to the more fundamental notion of the bad, and it is highly plausible that the relevant sense of badness must be a moral one—meaning that the person’s sense of the moral wrongness, too, must be dependent on this very type of badness. This is important because our sense of guilt is not usually linked with just any kind of values. For example, people do not usually report guilt for aesthetic failures. So, at the most fundamental level, we have some morally relevant disvalue. In cases of merely causal harm, most often, this disvalue seems to be pain, be it either physical or mental. Secondly, I have argued that in all the cases of causing harm without wrongdoing that the interpersonal theory usually refers to as evidence for their position, the harm is also perceived as noncomparatively unjust, which usually means that the harm is undeserved. So, even if one is not morally culpable for the relevant harm, they nevertheless view the harm as unfair towards the victim because the victim did not deserve to suffer such misfortune. If earlier, as the first layer of morality, I simply referred to the fact that cases of merely causal harm involve a reference to the same disvalue as cases of moral wrongdoing, then the aspect of unfairness gives ground to an even more special sense in which cases of causing harm are perceived as “not right”. This special sense of “not right” refers to the fact that the world is morally out of order—what people receive is not in accordance with their merit. Thirdly, there is usually also a further layer of morality that comes from the fact that the relevant desert base in cases of merely causal harm is almost exclusively that of one’s moral character. We are not concerned about harming people first and foremost because they are great mathematicians or rank highly by some other nonmoral talent, we view harm as undeserved because the person was a relatively decent person who did not deserve to suffer such a great misfortune or an especially good and kind-hearted person who should not even suffer lesser harms.

So, what might at first glance seem like a nonmoral instance of harm simpliciter, turns out to be a rather complex moral matter. In fact, it is quite amazing how many layers of morality can fit into a seemingly simple case. For those wondering by what trickery such a feat is possible, it may be useful to note that moral badness can come at different levels. Relying on Thomas Hurka’s recursive account of value (especially his remarks in (Hurka, 2001, p. 12)), we might label the pain that first catches the attention as the most fundamental base-level

disvalue involved in the case. As another lower-level<sup>76</sup> value, we have virtue (or at least sufficient decency), which remains somewhat more tacit and might not catch the eye at first. But desert definitely functions as a higher-level value here because the relevant desert judgment is purely a function of the combinations of the lower-level values/disvalues of virtue/vice and pleasure/pain. The combination of virtuous people suffering pain, in particular, is seen as bad. So, it is not that we have three different base-level values involved, the relevant values come at least at two levels, where the combination of lower-level values/disvalues like virtue/ vice and pleasure/pain gives rise to an additional layer of morality that consists in desert. It is because of this multi-level structure that there are so many layers of morality involved. Of course, in real life, we do not always explicate all those layers. We are so accustomed to taking them into account that we need not give any explicit thought to it. It is only when we take a closer look at cases of guilt for merely causal harm that all those layers are revealed. And even though this multi-level structure may sound complicated, it can actually simplify moral perception because, overall, we are operating with a smaller number of variables. But in any case, once we disentangle all those layers of moral concerns, there should be little doubt that guilt for merely causal harm can rightfully be called a moral emotion.

If this raises doubts about the overmoralisation of guilt, then let me end this chapter by revisiting the more general concerns about the overmoralisation of emotions and emphasising that the results of this chapter are products of sound reasoning that have not fallen prey to the common fallacies outlined at the beginning of this chapter, in section 3.1.2.

According to D'Arms and Jacobson (2000), our concern about feeling an emotion in a morally appropriate way can make us confuse moral appropriateness conditions with the general fittingness conditions of the emotion type. In other words, it can make us blind to the possibility that the emotion can be felt without reference to moral issues. Should this shed doubt on my analysis of guilt for merely causal harm as a matter of causing something morally bad? Is it the case that since we are so concerned with the morally bad, we become blind to the possibility that guilt can be felt about something that is not strictly morally bad but harm simpliciter? I do not think so because there is a clear sense in which those cases involve a reference to justice, which is definitely a moral matter. And the reason why I emphasise the importance of perceiving the harm as unfair is precisely because cases where harm is seen as fair do not seem to elicit guilt. As such, I have already taken care to be mindful of the moralistic fallacy, and

---

<sup>76</sup> Strictly speaking, I only mean that virtue is a lower-level value compared to the upcoming desert. It can, of course, be suggested that virtue itself derives its value from the fact that the subject must have the right attitude towards good and evil—the former they must love, while the latter they must hate. So, in this case, virtue would be a mid-level value. But I want to leave options open for those who attribute more primary importance for virtue, so this is why I do not argue that it must necessarily be seen as a mid-level value. In any case, in the text above, the most important point is that, regardless of how one attributes value to virtue, desert must nevertheless be seen as the most higher-level value involved in the cases of merely causal harm.

moralising the shape of guilt is clearly not due to the fact that I would have simply overlooked the possibility that guilt might concern simple badness rather than moral badness. The reason why I have moralised guilt is precisely because, when we look at the cases where guilt is reported, a nonmoral conception does not seem to suffice.

Next, is there a reason to suspect that emphasising the importance of moral badness somehow boosts the self-esteem of the person who feels guilt? I find this hard to believe. In case of the traditional account of guilt that tied guilt with moral wrongdoing, there was indeed the danger that viewing oneself as a moral agent capable of wrongdoing can boost one's self-esteem because it involves viewing oneself as someone who had control over what happened. However, according to my analysis of guilt for merely causal harm as a matter of causing a morally bad state of affairs, the locus of immorality is not the agent but the outcome. This completely removes the sense in which adding the element of morality could make the *agent* feel more powerful. There is, of course, some minimal sense of being powerful involved in being causally efficacious, but this is not relevant here because this sense of being powerful is not due to the moral aspect of guilt. Causal efficacy is equally present in case of causing harm simpliciter as well. One does not become more powerful by causing something that is morally bad. Let us imagine two parallel scenarios where an automobile driver kills another via an unfortunate accident. In one scenario, the person who died was a good person; in another, it turns out that they were a ruthless serial killer who had been a fugitive for quite some time. In one case, the driver feels severe guilt, whereas in the other, his initial guilt feelings are quickly dissolved because he views it as deserved punishment for the ruthless killer. Is there any sense in suggesting that viewing oneself as causing something morally bad involves viewing oneself causally more efficacious compared to the scenario where one just causes harm simpliciter? Is the first driver who killed an innocent bound to feel causally more efficacious than the one who killed a criminal? I do not think so. In both cases, their causal efficacy was equal—their driving their car caused another car to crash, and a person inside it to die. It is just the moral evaluation of the outcome that was different, not the driver's causal efficacy. Overall, it is hard for me to see that there would be any kind of self-boosting incentive to think of oneself as having caused something morally bad compared to having caused something that was not morally bad.<sup>77</sup> The fact that one has caused something morally bad only makes things worse. In that sense, the danger of overmoralisation seems even smaller when we conceptualise guilt as a matter of causing something morally bad compared to when we link it with moral wrongdoing that implies control.

So, overall, I conclude that we are perfectly justified in regarding guilt for harm to others as a moral matter, and that there is no reason to be concerned that this verdict is just a result of naïve overmoralisation.

---

<sup>77</sup> Perhaps someone thoroughly evil could experience a self-boost if they caused something morally bad, but I am here only talking about people who care about morality and would be susceptible to guilt. A person who feels guilt must necessarily view the morally bad outcome as something negative rather than positive.

## 4. SURVIVOR GUILT

This chapter is dedicated to an in-depth analysis of true guilt for unjust benefits, including the relevant type of survivor guilt. As such, I will not be discussing solidarity guilt that one may feel because they think they should share the fate of their loved ones in both good and bad. Neither will I be talking about the kind of guilt that one may feel because of the omission to work towards compensating for the injustices. Those alternatives, which can all be explained in terms of moral wrongdoing, were already set aside in section 2.2.4. The sole object of the kind of guilt that I am currently concerned with is the fact that one has an unjust privilege. Those cases are unique in that the subject need not have any kind of causal, let alone moral, responsibility for what happened to others. As such, it is especially unlikely that they should mistakenly construe themselves as having done something wrong. It is for this reason that cases of survivor guilt pose the greatest challenge to the traditional account of guilt, and push theorisation towards alternative conceptions.

As far as I know, what follows is the most detailed discussion of this particular type of survivor guilt to date. In general, thorough philosophical treatments of survivor guilt are relatively rare.<sup>78</sup> Usually, philosophers who write about guilt do not even make note of survivor guilt. For example, in a recent collection, “The Moral Psychology of Guilt” (Cokelet & Maley, 2019), survivor guilt is only mentioned on a handful of pages.<sup>79</sup> It seems to have become a stock example to dispute the traditional account of guilt, as if the issues surrounding survivor guilt were self-evident. Even if authors do pay attention to it, and even if they give it serious weight, they often confine their remarks to only a few pages (for a good example, see (Velleman, 2003, pp. 246–7)). And in the few articles that we do have about survivor guilt, the focus is not (or not only) on the kind of guilt that is solely felt for unjust benefits.<sup>80</sup> Most discussions of survivor guilt tend to fragment into explorations of various different types of guilt that a survivor may feel—including solidarity guilt, which (as shown in Chapter 2) is often confused with true guilt for unjust benefits.

---

<sup>78</sup> Some of the rare exceptions are (Fukuma, 2013; Griffioen, 2014; MacKenzie & Zhao, 2023; Metz, 2018).

<sup>79</sup> If this is to be read as criticism, then it should be considered as self-criticism as well because my own contribution to that volume included no mention of survivor guilt either.

<sup>80</sup> For example, Amber Griffioen (2014) talks about survivor guilt as just a form of traditional guilt that involves irrational beliefs, Satoshi Fukuma’s (2013) piece is more of a review of existing theories and seems to stop at something close to the already mentioned solidarity guilt, Thaddeus Metz (2018) who speaks of survivor guilt through the lens of African ethics seems to be emphasising solidarity as well. It was only very recently that Jordan MacKenzie and Michael Zhao (2023) published the most substantial article that offer us a the most thorough treatment of guilt for unjust benefits as such (though the article in general is equally dedicated to solidarity guilt as well). They discuss guilt for unfair benefits under the label of “luck guilt” (MacKenzie & Zhao, 2023, pp. 2717–2721).

So, this chapter focuses on guilt for unjust benefits as such, along with some of its developments. The last chapter was mainly dedicated to the question of whether guilt for merely causal harm is a moral matter. In the context of survivor guilt, we can skip this question because no one seems to doubt that the kind of justice that is involved in survivor guilt is a moral matter. Herbert Morris once wrote about survivor guilt in an article called ‘Nonmoral Guilt’ (1987) but his choice of terminology has been considered unfortunate because it allows deontic to “gobble up the entire domain of morality” (Murphy, 2012, p. 330), and even Morris himself suggests that guilt for unjust enrichment is derived from “a fundamental moral posture toward the world” (Morris, 1987, p. 222).

In this chapter, the central question is: how should we best capture the core relational theme of this kind of guilt? What is this guilt first and foremost about? It is easy to say that it has something to do with comparative injustice, but the real challenge lies in how exactly we should reframe the reference to injustice so that the rationale behind grouping survivor guilt together with more traditional instances of guilt would become more evident. What is it that justifies viewing survivor guilt as belonging to the same wider category as guilt for moral wrongdoing, and calling both by the same name? It cannot be that both must always have to do with comparative justice because we feel traditional guilt for going against various other values as well—for example, for harming others and thus having a negative effect on their well-being, or for violating their autonomy. At best, the mere reference to comparative injustice can grant a family resemblance, which links survivor guilt with only some forms of guilt for wrongdoing (cases where one is morally responsible for an unfair distribution of goods). This is indeed one way to explain the folk practice of extending the label of guilt to cases of unfair benefits, yet this would only go to show that the word ‘guilt’ is applied in a rather loose sense. The question of my interest, however, is whether there might be a more unified rationale for talking about guilt as a wider category that encompasses both survivor guilt and guilt for moral wrongdoing.

In the previous chapter, my strategy was to argue that the common explanation that reduces guilt for harm to others to harm simpliciter is too simple, and that there is much more to guilt for harm to others than first meets the eye. In this chapter, my strategy is, in a sense, the opposite. I will argue that most authors have tried to explain guilt for unfair benefits in a too roundabout way, failing to capture the most fundamental aspect—the injustice of the situation as a morally bad state of affairs. The focus is always put somewhere else: Roy Baumeister and his colleagues (1994) emphasise the similarity with harm to others, suggesting that in both cases, the subject must feel that the privilege comes at the expense of other people. David Velleman (2003) says that guilt involves a sense of normative vulnerability, which manifests itself in anxiety about warranted envious resentment (imaginatively attributed to those who did not survive, or their loved ones who feel envious resentment on their behalf). Jordan MacKenzie and Michael Zhao (2023), in great part inspired by Velleman’s idea of normative vulnerability, argue that the key element consists of a perceived lack of self-justification. Even though I do not deny that all those explanations can illuminate and give us a fuller

account of many instances where guilt is felt over unjust benefits, I will argue that they fail to give due weight to the most simple and fundamental element that is responsible for many of those further symptoms (sense of vulnerability, lack of self-justification, anxiety about warranted resentment) and the identification of which is key to integrating inequity guilt into a more general theory of guilt. This most fundamental underlying element lies in the subject's perception of themselves as a constitutive element of a comparatively unjust—and hence also, a morally bad—state of affairs.

#### **4.1. At-the-expense guilt and harm to others**

I will begin with how the interpersonal account of guilt, which views guilt as a matter of causing harm to others, could try to explain survivor guilt. Before I move on to Baumeister's own suggestion, though, I would first like to review how one might try to make sense of guilt for unfair benefits in terms of literally causing harm—just to illustrate why direct causing of harm does not really offer a particularly promising strategy to illuminate the type of survivor guilt that we are concerned with here—inequity guilt for unjust benefits. We can get a clue for such a strategy from Gabriele Taylor (1985). Taylor does not think that guilt is necessarily a matter of causing harm and prefers to speak in terms of causing some forbidden state of affairs, but this need not concern us here, because the problem arises due to the causal element that is common to both views. The way that Taylor has tried to explain inequity guilt in the context of social status is the following: “If I feel guilty about my privileged position in society due to circumstances of 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” (Taylor, 1985, p. 91).<sup>81</sup> Patricia Greenspan calls this comment “puzzling” and points out that one would have to be deluded to an implausible extent in order to think that one caused their birth (Greenspan, 1992, p. 297, 1995, p. 231). To be fair, I suspect that what Taylor has in mind is not necessarily the idea that one's birth was one's own doing. Remember the cases of being thrown on top of a pile of bodies—one may be passive in those cases as well, but nevertheless perceive themselves (qua physical body) as part of the causal chain that leads to harm to others, and thus feel guilt. It is definitely true that one's birth is a necessary causal precondition of the current socially unjust state of affairs that one perceives oneself to be part of. That being said, it does still seem bizarre to track one's guilt for privileged social status to something as far removed as one's birth—one's birth simply does not seem like a particularly salient feature in light of which we should construe the current injustice. There is, of course, a sense in which it is in itself unfair that some people are born into wealth and privilege, while others are

---

<sup>81</sup> Even though Taylor is not technically a proponent of the interpersonal theory of guilt, her view is nevertheless close in the sense that she emphasised causal responsibility for a prohibited state of affairs.

born into poverty, but this is not what Taylor is saying—she is suggesting that one’s birth is the cause of the current unjust state of affairs. Even though one’s birth is indeed a *causal precondition* of (at least part of) the current injustice, it seems farfetched to construe it as a *cause* of the injustice because it does not cause others to be born into worse situations. Whereas in the non-agential case of being thrown on top of others, there is quite a clear sense in which one’s body causes pain to others.<sup>82</sup> The case of social privilege is, of course, just one example of guilt for unfair benefits, but it is also hard to see how such a strategy might help us make sense of the more particular form of guilt for unfair benefits that we can see in the case of survivor guilt. In those cases of survivor guilt that I am currently concerned with, which are indeed true instances of guilt for unfair benefits, there simply does not seem to be a relevant sense in which one causes the injustice of them having survived while others did not. There is, of course, a sense, in which your existence is caused by your birth, so birth does function as a precondition (of, really, anything that happens to us), but it does not make sense to construe one’s birth as a cause of the unjust state of affairs that consists in one having survived while others did not. For those reasons, I do not think that guilt for unfair benefits can be explained in terms of directly causing harm to others.

Next, I will move on to Baumeister’s own suggestion with regard to how we should construe inequity guilt in terms of harm to others, which is a bit more roundabout, yet more plausible than Taylor’s. The interpersonal account of guilt explains all kinds of guilt via their resemblance to guilt for harm to others, and the main point of similarity that Baumeister and colleagues emphasise is the fact that the subject benefits at the expense of the victim:

Positive inequity may resemble the prototype causation of guilt in the sense that when one harms a relationship partner, one presumably benefits at the partner’s expense; thus, benefiting inequitably—and thus seemingly at the partner’s expense—may be sufficiently similar as to cause guilt feelings. (Baumeister et al., 1994, p. 247)

So, Baumeister and colleagues suggest that the eliciting conditions of survivor guilt must be similar to the prototypical scenario of causing harm (which includes cases of wrongdoing): in both cases, the subject benefits, at least seemingly, at the other person’s expense. I will call this type of survivor guilt at-the-expense guilt. First, I will elaborate further on what kind of guilt it is and which situations it captures. Even though I do not think that it is a suitable explanation for all guilt

---

<sup>82</sup> Of course, it might be suggested that even in the latter case there is a sense in which one is not the final cause because they are thrown by other people, but I am not concerned with final causes here. I am concerned with whether one functions as a cause to the extent that, if we removed them from the relevant causal chain of events, would the end result have involved less harm to the victim. And this does indeed seem to be the case—if we remove the person who is thrown on top of others, then others would not get crushed. This is not the case with social privilege—there is no obvious sense in which one not being born makes things better for those who are born into poverty.

for unfair benefits, it is nevertheless an important type of guilt that captures the experience of many survivors. Eventually, however, I will argue that not all inequity guilt takes the form of at-the-expense guilt.

At-the-expense guilt involves a sense that one's good fate was in some way dependent on other people's bad fate, that one benefitted from the suffering of others. Baumeister's reference to the fact that one's good fortune is perceived as coming at the expense of those who were less fortunate does indeed capture an important feature of the so-to-say original survivor guilt—that is, survivor guilt as it was often described in the context of Holocaust survivors, back when the term 'survivor guilt' was coined by William G. Niederland (1961). This was a context where resources and other goods were incredibly limited, which meant that if one got more, then the other got less. So, whatever privilege one enjoyed, this was likely to come at the expense of other people.

To get it out of the way, let us first acknowledge that much of the original at-the-expense guilt took the form of traditional guilt. For example, one might have cheated during the elections. People used egoistic and sometimes perhaps even unfair means to stay on top of the competition and not to be chosen during the selections (as those who were meant to be sent to the gas chambers). This included disguising one's emaciation by stuffing cloth into their clothes or mouths (Lifton, 1991, p. 490), but also means employed to stay less sick and malnourished, which included not sharing your food with others. Those kinds of reasons are also reflected in the words of the Holocaust survivor Bruno Bettelheim, who makes it especially clear that staying strong came at the expense of one's fellow prisoners:

[T]he voice of the conscience replies: "True, but the reason you had the chance to survive was that some other prisoner died in your stead." And behind this in a whisper might be heard an even more severe, critical accusation: "Some of them died because you pushed them out of an easier place of work; others because you did not give them some help, such as food, that you might possibly have been able to do without." (Bettelheim, 1979, p. 27)

There are reports of prisoners altering the lists of people who were meant to be sent to the gas chamber, and replacing their own names with those of other inmates (Lifton, 1991, p. 490). Some also hid themselves for the selections (Modell, 1971, p. 341). All those cases may give rise to a rather traditional form of guilt that is felt for one's own actions.<sup>83</sup>

But at-the-expense guilt felt by Holocaust survivors is definitely not exhausted by cases where one may be said that have engaged in morally suspicious means in order to avert death. Even if one did not attempt to cheat at the selections (and does not see themselves as having done anything morally wrong), there is nevertheless a sense in which their staying alive came at the expense of others. Most

---

<sup>83</sup> In those traditional cases of at-the-expense guilt, the at-the-expense part can also be described in terms of the metaphor of stealing. In Primo Levi's (1988) words: "each of us [...] has usurped his neighbor's place and lived in his stead" (p. 57).

importantly, it should not be forgotten that, in concentration camps, death quotas were indeed a real thing—the number of people who were supposed to be sent to the gas chamber was sometimes indeed fixed (Modell, 1971, p. 341).<sup>84</sup> So, if one was not chosen for extermination, this meant that someone else had to go because the numbers had to be fulfilled. It was because others were sent to the gas chamber before, that one managed to stay alive. Also, in addition to death quotas that pertained to direct extermination, when prisoners were transported from one location to another, the railroad cars were so overloaded that it was inevitable that some prisoners had to die because they were pushed and stamped on. To put it in more numerical terms: if a hundred people were loaded into a car that was meant for forty and even if everyone was standing had room for no more than eighty, then it was inevitable that twenty people in each wagon had to die. (Wiesel 2006, 99; see also Lifton 1991, 491) Even if one was not themselves stamping on other people (a case that might give rise to regular guilt), there is nevertheless a sense in which their staying alive came at the expense of those who got stamped on. The mere fact that they were there in the wagon, standing up and taking up precious space, meant that someone else had less space, and ended up being stamped on. In all those cases, there is no irrationality involved in saying that one's survival came at the expense of other people.

The experience of at-the-expense guilt is also not limited to the context of the Holocaust. The kind of situation where the number of people who can stay alive is limited can also be found in the context of modern medicine. Let us consider a case where one received life-sustaining treatment at the expense of another patient. It is not one's own doing that he is saved and not the others, but it is nevertheless a burden to be involved with injustice in such a way. Some interesting cases of this guilt, especially in people who have received an organ transplant, can be found from (Blacher, 2000). And neither does at-the-expense guilt necessarily have to be a matter of life and death. In the context where a company is just looking for one new employee, the one who gets the offer may feel that he received it at the expense of other candidates, who may have been more qualified.

And in addition to rational cases of at-the-expense guilt, there are also cases that rest on irrational beliefs about scarcity. The mechanisms of this fallacy have been described by Arnold H. Modell as follows:

There is, I believe, in mental life, something that might be termed an unconscious bookkeeping system, i.e., a system that takes account of the distribution of the available 'good' within a given nuclear family so that the current fate of other family members will determine how much 'good' one possesses. If fate has dealt harshly with other members of the family, the survivor may experience guilt, as he has obtained more than his share of the 'good'. (Modell, 1971, p. 340)

---

<sup>84</sup> Of course, one might irrationally think that there are death quotas in natural catastrophes as well, that there is a certain number of people who are statistically likely to die and that if one remained alive then this meant that someone else had to die. But here, I will only focus on cases where one can rationally say that one survived at the expense of someone else.

He is suggesting that people think of good fortune within a nuclear family as a limited resource, so that if one gets more of it, others will necessarily get less.<sup>85</sup> Even if such beliefs are irrational, the kind of guilt they yield is nevertheless the same kind of at-the-expense guilt that does not require one to think that they themselves did something morally wrong.

At-the-expense guilt captures many instances of survivor guilt, especially in the context of the Holocaust, which gave rise to much of the original literature on survivor guilt. So, it is understandable why discussions of this aspect are emphasised in so many classic texts. However, when we look at the wide range of cases where the term survivor guilt is applied today, not all of them fit this pattern. Appealing to benefitting at the others' expense does not seem like a particularly good explanation for the kinds of survivor guilt that a cancer survivor may feel about the unfairness of the fact that they made it out alive while the patient with whom they shared the same hospital room did not. Or imagine a case where one sibling feels guilt in front of the other because he was born healthy while the other was born with a severe disability. To suggest that, in those cases, one must think that their good fortune comes at the expense of others would be nonsensical. There is often simply no reason to attribute to people the kind of belief in a moral bookkeeping system that Modell described. We don't normally assume there's a fixed quota for how many people can survive cancer, or for how many can be born healthy and without disability. There is, of course, statistics—we may know that a certain percentage of people are likely to recover, and a certain percentage of children are born healthy—, but this is not the same as the Holocaust death quotas, and neither do people necessarily have to think about the two in the same manner. Holocaust death quotas were set before the executions—they set the norm, and the executions just followed the plan. Statistics, however, is something that we do post-fact—it simply describes what has happened in the past; it does not set a norm for what must happen in the future. It is, of course, not impossible that some people may indeed irrationally think of statistics as setting a norm. Yet I find it implausible to suggest that this is how all the relevant survivors must necessarily construe the situation. But the explanation of survivor guilt that Baumeister offers rests on assuming that people must always have such beliefs, and thus much of survivor guilt would have to be unfitting—tied to irrational beliefs about life being a zero-sum game where one's good fortune necessarily comes at the expense of other people's good fortune.

---

<sup>85</sup> In fact, Michael Friedman (1985) has labelled this kind of guilt 'depletion guilt' (p. 26): since one's good fortune comes at the expense of others, it seems as if they would have depleted to source of good fortune, so that there is nothing left for others. The reason why I am reluctant to call it depletion guilt is that the term seems to imply that it was the subject who depleted the resource and left nothing for the others, thereby putting the responsibility on the subject. While this is how people may sometimes perceive at-the-expense guilt, I do not think that such agential perception is a necessary element of the at-the-expense guilt.

So, even though the interpersonal account, at least as originally proposed by Baumeister, does not reduce survivor guilt to a matter of perceived wrongdoing, it will nevertheless end up having to attribute irrational beliefs to the survivors—only those beliefs will concern one surviving at the expense of others. While I do not deny that there may be cases where irrational beliefs are involved, I find it implausible to suggest that all people who report survivor guilt must have such irrational beliefs. This is similar to how cases of merely causal harm, too, may sometimes involve irrational beliefs about being morally responsible but need not do so. Yet, while the interpersonal theory vindicates all cases of merely causing harm as perfectly rational ones, survivor guilt is often turned into an irrational reaction. It is not at all obvious to me that this should be the case. It seems that by putting the focus on causing harm to others, the interpersonal account of guilt has prioritised cases of merely causal harm at the expense of guilt for unfair benefits and the corresponding type of survivor guilt. The solution that I will eventually defend, partly at the end of this chapter, but in greater detail at the beginning of the next one, enables us to view both types of cases—cases of merely causal harm and survivor guilt as standing on par, both having the potential to be, at least in some cases, perfectly fitting and rational responses.

## **4.2. Normative vulnerability and anxiety about resentment**

If the interpersonal account does not enable us to do justice to survivor guilt, where should we turn to instead? Once we have decided that the traditional account that reduced all guilt to perceived wrongdoing is too narrow, there are not that many alternatives to the interpersonal account of guilt. There is, however, one interesting line of thought that was originally defended by David Velleman (2003) and has recently been revived by Jordan MacKenzie and Michael Zhao (2023). This line of thought builds on the notion of normative vulnerability, which means “the sense of being unjustified, of having nothing to say for oneself” (Velleman, 2003, p. 235). Velleman himself ties the sense of normative vulnerability to anxiety about warranted negative responses like loss of trust and resentment. According to his analysis, survivor guilt resembles traditional guilt because both involve anxiety about warranted resentment. While MacKenzie and Zhao do not explicitly emphasise the aspect of anxiety, they nevertheless build on Velleman’s notion of normative vulnerability and stress one’s concern for self-justification, which only makes sense in the background of anticipating negative responses from other people. In what follows, I will review and argue against both views. While I do not deny that many instances of guilt may be accompanied by Vellemanian elements, I am fundamentally opposed to the idea that we should explain guilt in terms of anxiety about negative reactions towards us and concerns about self-justification. The root of the problem lies in the fact that Velleman builds on a Freudian view of guilt, which puts a great emphasis on fear of punishment. As I will argue, such a view is fundamentally flawed because it makes guilt into a matter of egoistic motivation. And once we purify the

Vellemanian account from all the Freudian deadweight, i.e. element of anxiety and vulnerability, we can see that most of the unification work is actually done by the notion of justice as a form of moral badness—which is what we should have started with in the first place.

#### 4.2.1. Velleman on guilt as anxiety about resentment

Let us begin with Velleman's original view, as defended in his article entitled "Don't Worry, Feel Guilty" (2003). The title is, of course, ironic, because guilt itself involves a fair deal of worry, but the idea is that we should not be worried about feeling guilty in the absence of wrongdoing because guilt can be a rational response even if one has done nothing morally wrong. As such, his overall goal is very similar to the dissertation at hand. But while I seek to accommodate guilt for merely causal harm and survivor guilt, Velleman is concerned with self-disciplinary guilt and survivor guilt. In order to do justice to both, he ends up defending "a disjunctive analysis" (family resemblance view) of guilt (p. 236), according to which guilt consists of anxiety about having warranted some kind of negative response like distrust and resentment (p. 247). Whereas traditional guilt can consist of anxiety about both distrust and resentment, self-disciplinary guilt is similar to traditional guilt because it involves anxiety about distrust, and survivor guilt is similar to traditional guilt because it involves anxiety about resentment. As I have already argued in 2.2.2, I do not think that self-disciplinary guilt poses a particularly serious problem to the standard account of guilt. For this reason, I will not be concerned with that part of Velleman's disjunctive analysis that focuses on self-disciplinary guilt and anxiety about loss of trust. My interest lies in how he tries to unify the sphere of traditional guilt and survivor guilt by saying that both consist of anxiety about warranted resentment, and his more general ideas about guilt as a matter of normative vulnerability.

Let us begin with the latter—the concept of normative vulnerability—, because this is the foundation on which Velleman builds his account of guilt. At the heart of normative vulnerability is "a sense of being somehow unjustified, of having nothing to say for oneself" (Velleman, 2003, p. 235). As such, it is not defencelessness against a bully who has no right to treat the subject in a harsh manner. It is a sense of *normative* defencelessness—one feels defenceless against negative responses (like that of resentment) in the sense of having no claim to be spared from them because those responses are perceived as warranted (fitting). (pp. 235–236). As Velleman goes on to explain, the idea of normative vulnerability opens the key to understanding the dual nature of guilt:

The concept of normative vulnerability helps to explain why guilt is a feeling of both anxiety and diminished self-worth. The anxiety comes from feeling oneself exposed to something untoward. The sense of diminished self-worth comes from conceiving of that exposure as a matter of being stripped of a claim or entitlement. (Velleman, 2003, pp. 235–236)

By emphasising that guilt is a feeling of both anxiety and diminished self-worth, Velleman paints a rather unique picture of guilt. He makes guilt a Janus-faced emotion. Let me explain. Janus was the Roman god of beginnings and endings, transitions and doorways, who is depicted as having two faces—one looking at the past, and the other looking at the future. According to Velleman, guilt, too, has two faces—one that looks at the past, at what has happened and how this has somehow put the subject in the wrong and stripped them of entitlement; and the other that looks at the future, at the negative responses that one may have to face. Later on, Velleman does not really elaborate on the aspect of diminished self-worth and is more concerned with the anxiety part, but this is the more complete architecture that he begins with, which will become important later on in my analysis.

Let us currently keep the focus on anxiety, though. After all, for the most part, Velleman himself, too, speaks of guilt in terms of anxiety. In general, anxiety is understood as an emotion that we experience when we perceive some uncertain threat in the future. As such, it differs from fear, which is experienced in the face of imminent threats. (Kurth, 2018, p. 46) So, for example, we can talk about fear when we are currently at the mercy of another person's violent anger. But when we are worried about what kind of sanctions our moral transgressions may bring about in the future, it is more precise to speak in terms of anxiety rather than fear. Since we do not know yet what exactly will happen, the object of anxiety is more diffused.

The idea that guilt is, at the bottom, a form of anxiety reflects the fact that Velleman works within a Freudian framework. He starts from a Freudian analysis of guilt as anxiety about the prospect of being punished by an internalised authority figure, the super-ego,<sup>86</sup> and sets out to improve upon this account by adding a normative element that emphasises that guilt is not just “fear of a bully” but anxiety about *warranted* negative responses, and further clarifying what those negative responses are because it is not quite obvious that we are justified to outright punish people for private wrongs like breaking a promise or hurting someone's feelings (Velleman, 2003, pp. 236–237). As such, he seeks to offer “a rationalist revision of Freud's analysis”, according to which “guilt is anxiety about having warranted treatment that is sometimes formalised as punishment” (p. 244).

The type of negative reactive attitude that Velleman thinks is warranted in the context of traditional guilt and survivor guilt is resentment. While in traditional instances of guilt, we experience anxiety about being on the receiving end of angry resentment, survivor guilt is supposed to consist of anxiety about envious resentment. Anger and envy are, of course, different emotions, but to the extent that both are forms of resentment, they form a “natural pair” that is unified by the fact that both involve an embittered sense of injustice (Velleman, 2003, p. 247).

---

<sup>86</sup> See, for example, Freud's definition of guilt in “Civilization and Its Discontents”: “the sense of guilt is at bottom nothing else but a topographical variety of anxiety; in its later phases it coincides completely with fear of the super-ego” (Freud, 1962, p. 82).

Indeed, a situation where one has done something morally wrong or has been the recipient of an unfair privilege can often give rise to anxiety about resentment. In cases of traditional guilt where we have transgressed against another person, it is quite likely that they will respond with an adverse reaction that involves a sense of being unfairly treated. While it might not be obvious at first glance that survivors are worried about being the object of envy, there is a case to be made for that as well. A simple reason why one might be sceptical about survivors being anxious about being on the receiving end of envious resentment is that the others toward whom it is comparatively unjust that the subject made it out alive are dead and thus no longer in a position to resent one. In response to this objection, Velleman (2003, p. 247) suggests that third parties can nevertheless feel resentment on behalf of the deceased. This may indeed be the case, especially when we think about the loved ones of the deceased. There are examples of how survivors seek something akin to forgiveness from the loved ones of those who did not make it out alive. For example, a documentary called “Sole Survivor” (2013) tells the story of George Lamson, who was a sole survivor of a plane crash, and later visits the families of the deceased in order to seek forgiveness.

But, in addition to the fact that anxiety about resentment may turn into anxiety about the resentment that others may feel on behalf of the deceased, there are also many reasons why survivor guilt can nevertheless reflect anxiety about the resentment of the deceased themselves as well. Even though there is no danger of being faced with the actual resentment of the deceased, one may nevertheless be confronted by them in their imagination and dreams. For example, such dreams are reported by a former United States Army medic, Adam Linehan, who served in Iraq and Afghanistan:

I also started having frequent dreams about the apocalypse. Sometimes soldiers I knew appeared in them, one more consistently than the others. It was always the same scenario: We are standing face to face, and I am trying to explain something to him, I’m never sure what, and he just stares at me blankly, as if he’s annoyed or disappointed, or as if I’m making excuses, dancing around the point, and his impatience is an unnecessary reminder that I’m living on borrowed time. (Linehan, 2019)

Furthermore, survivor's guilt (or inequity guilt in general) is not only felt with respect to those who are dead. When writing about survivor guilt in the context of the atom bomb experience in Hiroshima, Robert Jay Lifton (1991, p. 52) notes that those who were severely burned or otherwise injured did indeed harbour a fair deal of resentment against those who were not. So, it is only to be expected that those who were unscratched were anxious about being confronted with this resentment. And one can also feel inequity guilt in relation to the dying. For example, Lifton (1991) mentions how a doctor who attended dying patients recounted that he had to “contend with the guilt evoked by the questioning glance of the dying, with the unspoken question, ‘Why should I die while you live?’” (pp. 491–492). In fact, it seems a certain type of glance is especially powerful in

communicating this kind of resentful sense of injustice, which is probably the reason why it appears in dreams as well (remember that Linehan mentioned a similar kind of disappointed stare as well).

But does the fact that the scenario that elicits guilt can also elicit anxiety about resentment, necessarily imply that guilt itself should be identified with this type of anxiety? Or with anxiety in general? Regardless of all the sophisticated improvements that Velleman makes to the Freudian account of guilt, the main problem with his account arises from the fact that it is built on the Freudian foundation that assimilates guilt with anxiety. In what follows, I will rehearse a number of counterarguments against this more general idea that guilt is a type of anxiety but let me preface this by noting that there is a sense in which this is a useless undertaking. I take it that for most of my readers, it is obvious that guilt and anxiety are two completely different emotions, so they will not need much convincing at all. For example, Baumeister does away with fear-based accounts of guilt in just a few passing sentences, noting that the two emotions have different intentional objects and suggesting without any further examples that “[o]ne can clearly feel guilt in cases in which there is no likelihood of punishment and hence no fear” (Baumeister et al., 1994, p. 245). Yet I take it that for others, who are already deeply drenched within the Freudian framework, none of my arguments will do because it would require a complete paradigm shift to abandon the anxiety-based account. But let us first rehearse the arguments, and then come back to the disagreements that may remain.

The simplest and the most intuitive, though not necessarily conclusive, strategy to demonstrate that guilt is not the same as anxiety about some negative sanctions is to give examples of cases where guilt comes apart from anxiety. While I will have more to say about the more substantial reasons why guilt is not anxiety later on, let us first focus on the cases because this will help us stay in contact with how the issues of guilt and anxiety are exemplified in real life. I will confine myself to highlighting two examples that I myself have found particularly compelling. The first comes from Michael Friedman (1985b, pp. 532–533) and concerns Miss L., who was the only child of an elderly couple who decided to move out of her parents’ home at the age of 30. Her parents loved her dearly and recognised the importance of her independence. She was not anxious about having to face negative sanctions like resentment or loss of love. What she felt guilt about was the mere fact of harming them, of making them feel sad and alone when moving out. It was not anxiety about sanctions that might have followed.

Secondly, let us consider a case from Janette Turner Hospital’s short story “Night Train” (2003), which has also been highlighted in Christopher Cordner’s (2007) critique of anxiety-based construal of guilt. When in fifth grade, Brian felt sorry for another boy, Albert, who was regularly bullied by others at school. While Albert is being bullied, Brian does not help him or defend him, sometimes he even laughs with the other boys:

Sometimes I laughed. I always felt sick when I did. The worst thing was, he knew that. He would give me this sad little smile. He understood. He forgave me. I'd go outside and I'd want to hang myself, like Judas. I used to vomit in those bushes outside the school. (Turner Hospital, 2003, pp. 242–243)

Brian obviously feels an intense guilt at not defending Albert from bullies. However, there is no prospect of resentment or distrust from Albert—he understands and forgives. Brian knows that he is not exposed to any negative sanctions, given Albert's understanding and forgiveness, and that makes him feel even more guilty for betraying such a kind person.

Both cases exemplify a scenario where one feels guilt about their actions with respect to someone who loves them nevertheless and will not hold it against them. Technically, one could, of course, always insist that there is some unconscious anxiety, but this would turn the claim into an unfalsifiable dogma. Rather than giving ground to the idea that there must have been some unconscious anxiety, those cases seem to suggest that one's sense of guilt may be even stronger when they do not anticipate any negative reactions because it feels especially unfair to hurt (or not help) someone who is so loving and kind towards you.

Those cases showed that anxiety is, in general, not necessary for guilt. But what about the necessity of anxiety about resentment in the context of survivor guilt in particular? To begin with, it seems that we can think of similar counterexamples. For example, in Auschwitz, once one prisoner had successfully escaped, ten prisoners were chosen to be sent to death by starvation as a punishment. However, a priest Maximilian Kolbe volunteered to take the place of a fellow man who had a wife and children. (Pojman & Fieser, 2012, p. 149) Imagine what the man might have felt. I find it plausible that he could have felt survivor guilt because his life was quite literally saved at the expense of Kolbe's,<sup>87</sup> but it is especially unlikely that he must have been worried about resentment on Kolbe's behalf because Kolbe was a virtuous man who himself volunteered to take his place. Again, we have an example of a victim who is especially loving and kind.

But, of course, none of what I have said rules out the possibility that one might be anxious about the resentment of (real or imagined) third parties. I personally find this implausible, and I think that at least the case that involved moving out of one's parents' house should have conclusively teased guilt apart from any kind of anxiety, even on behalf of third parties—because there is nothing wrong with a 30-year-old woman starting an independent life.

But in the case of survivor guilt, I can nevertheless see a potential line of argument that one might try to push in order to defend the idea that survivor guilt is a matter of anxiety about envious resentment. Also, let us not forget that it is in great part precisely in order to do justice to survivor guilt that the Vellemanian approach has emerged. So even if it would not work as a general account of guilt, it may nevertheless provide us with useful insights into the survivor experience.

---

<sup>87</sup> Even though, as I have previously argued, feeling that one survived at the expense of other people is not necessary for survivor guilt, it is definitely an intensifying factor.

The line of thought that I have in mind seems to be expressed, albeit somewhat vaguely, by MacKenzie and Zhao (2023), who have recently attempted to revive the Vellemanian account of guilt by placing a great emphasis on the idea of normative vulnerability—the sense of having nothing to say in defence of oneself. They suggest that such a perceived lack of self-justification resonates especially well with the fact that survivors are often haunted by the “Why me?” question, which is not simply a demand for a causal explanation but a call for normative self-justification (p. 2717). It is tempting to interpret them as suggesting that the “Why me?” question, that has become almost definitive of survivor guilt, is a demand for self-justification that mirrors how those who died might confront the subject with the kind of envious resentment that Velleman earlier spoke about: “Why you? What do you have to say in defence of yourself?”. At least this is what seems to be implied by the scenario that they suggest: “guilt arises in these cases because the survivors recognise that there is no satisfactory answer to the justificatory [‘why me and not them?’] question” (p. 2717). So, the idea seems to be that one confronts oneself with an accusatory demand for self-justification, realises that their good fortune is indeed unjustified in the sense of being unfair, and develops feelings of guilt. Later on, when the question gets repeated many times over, it need not even be a genuine question anymore but can simply become a rhetorical way of saying to oneself in a resentful manner “Look at you being alive and well! How unfair!”. This is indeed one way we can interpret the “Why me?” question, and there is definitely some textual evidence to support that this is how survivors construe it. For example, when Robert Lifton talks about the survivor guilt of the atomic bomb survivors in Hiroshima, he says: “Everything returns the *hibakusha* [survivor] to his self-accusatory inner question: “Why did I live while they died?”” (Lifton, 1991, p. 55) In fact, I suspect that to some extent at least, MacKenzie and Zhao have based their interpretation of the “Why me?” question precisely on their reading of Lifton because they also cite Lifton to back up their claim that guilt is a matter of perceived lack of self-justification: “From the moment of the atomic bomb exposure, the *hibakusha* [survivor] experienced a need to justify his own survival in the face of other’s deaths” (Lifton 1991, 35; cited in MacKenzie and Zhao 2023b, 2709). If the “Why me?” question is indeed an integral part of survivor guilt, and if it is correct to interpret this question as a resentful demand for self-justification (or a resentfully implied lack of self-justification), then one could indeed make a case that survivor guilt can be modelled as anxiety about envious resentment because one would be especially concerned and worried about the scenario of being faced with resentful attitudes and not being able to justify oneself.

Perhaps this is indeed sometimes the case. But it seems to me that many self-reports tell a different story. I will argue that, in those cases at least, the above-mentioned construal misrepresents, first, what the “Why me?” question is ultimately about; secondly, the right sense in which the survivors find that there is no satisfactory answer to the question; and thirdly also the temporal unfolding of the overall “Why me?” experience.

I will not doubt the prevalence of the “Why me?” question among those who feel survivor guilt, but I do think that we should be careful about how to interpret the meaning of this question. In particular, I am worried that if we approach it from a secular normative ethics perspective, this question can become something other than what many survivors actually have in mind when asking it. Let me suggest an alternative story which does not model the question as a resentful confrontation. I will argue that, in many cases, the abovementioned construal misrepresents, first, what the “Why me?” question is ultimately about; secondly, the right sense in which the survivors find that there is no satisfactory answer to the question; and thirdly also the temporal unfolding of the overall “Why me?” experience.

Let us first explain how else we might interpret the “Why me?”, and in what tone it is asked. For this alternative interpretation, we get an important clue from Bruno Bettelheim. When he, as both a psychologist and also a Holocaust survivor himself, discusses the “Why me?” question, he notes: “It was a miracle that the survivor was saved when millions just like him perished, so it seems that it must have happened for some unfathomable purpose” (Bettelheim, 1979, p. 27). This puts the “Why me?” question in a completely different light. The question is not posed as an accusatory demand to justify oneself. It arises as an attempt to understand the higher purpose behind their survival, an attempt to understand God’s ways. As such, it does not arise first and foremost as an accusatory demand but reflects a desire to resolve the conflict between the seeming unfairness of the situation and one’s deeply held belief that the world is ultimately a just place where everything happens for a reason. This is why the “Why me?” question has been described as part of the “on-going dynamic of making sense of why they survived” (Pethania et al., 2018, p. 1). As put by one of the informants in a study by Yasmin Pethania and colleagues: “*I was brought up in a religious upbringing so they talked about God and good and evil, those things. When I started looking at death in, it’s actual, vivid horrible, then nothing makes sense anymore what does all this mean?*” (2018, p. 3) It is clear that the quest for sense-making comes from the clash between one’s religious belief in a just world and the seeming injustice of death.

Secondly, framing the “Why me?” as an attempt to understand God’s ways helps us better understand the sense in which survivors often find that the question has “no satisfactory answer”. It is not that the survivor must necessarily think that their survival is unjustified and unfair (as MacKenzie and Zhao (2023, p. 2717) seemed to imply when they suggested that the “Why me?” question be viewed in connection to a sense of being unjustified). In the latter case, there is nevertheless a sense of finality in realising that there was no justice in the subject outliving others who were equally worthy. There is an answer, just not a satisfying one in the sense that it does not speak in the subject’s favour. But survivors seem to characterise the “Why me?” question as posing an “unsolvable riddle” (Bettelheim, 1979, p. 26). I find it plausible that this just reflects the idea that it is beyond one’s capacity to know the answer. This seems evident when we look at a more complete account of a case mentioned by MacKenzie and Zhao as well. It is the

self-report of Jerry Schemmel, a sportscaster who survived a plane crash that killed over 100 passengers.

It was a long time after the crash, a couple of years, where I asked [myself] the question every day - why did I survive the crash? Why did I survive and this little boy in front of me die? Why did this guy beside me die and [my friend] Jay seven rows behind me die? Why? Why?

When I finally came to the point where I realised I would never have those answers, it would never make sense, I stopped asking the question so often. That's when I think I got a lot healthier.

I was like 'Alright, I'm going to just leave this in god's hands and I'm going to let him deal with it. I'm here, I'm happy I survived, I'm going to make the most of it.'<sup>88</sup>

As we can see, he realises that he “would never have those answers” and just decides to trust God and leave the matter in his hands.

Thirdly, this also reveals a rather different temporal unfolding of the “Why me?” experience. According to the model that seemed to be suggested by MacKenzie and Zhao, one started off with the “Why me?” question, then realised that it is indeed unfair that they survived and then developed feelings of guilt. However, when looking at the case of Schemmel, we can see how realising that there is no satisfying answer is the end, rather than the beginning of guilt—or at least the beginning of an end, because this is what ultimately helps him overcome his guilt and get better.

But if this is the case, then we might ask: at which point with regard to the “Why me?” question does guilt arise then? My suggestion is that it arises in the very beginning, or perhaps it is even more correct to say that this is what gives rise to the “Why me?” question in the first place. The subject already starts off from the recognition that, at least *pro tanto*, the fact that he outlived others seems unfair—if not unfair in the all-in-all sense of the word, then at least with respect to those others, and this is already mirrored in a sense of guilt. However, since the person has a background belief in a just world, they cannot just accept that the world is an all-in-all unfair place—they would like to think that there must be some higher purpose behind their survival, and the question “Why me?” seeks an answer to what it is that makes one's survival, even though comparatively unfair towards others, all-in-all nevertheless a good thing. The survivor wants to understand what their mission in the world is—the mission that is supposed to justify the fact that they were spared. They want to know with which plan in mind did God choose to spare them.

This is similar to how many Jews who were confronted with the atrocities of the Holocaust, and felt the emotional reaction of deep injustice, pondered about the mysterious ways of God and raised question about “Where is merciful God?”

---

<sup>88</sup> “‘Everybody Around Me Died’: Jerry Schemmel’s United 232 Survivors’ Guilt.” <https://www.offtheball.com/other-sports/jerry-schemmel-survivors-guilt-1048629>

(Wiesel, 2006, p. 64) and “Why on earth is this happening?”. Here, too, the question about why all this is happening is already preceded by a strong emotional reaction that reflects a deep sense of injustice.

When thinking about survivor guilt today, and especially in the context of more secular disciplines, it may be easy to forget that much of the original survivor guilt that was felt by the Jews who survived the Holocaust, was the kind of guilt felt by deeply religious people.

As I have argued, the “Why me?” question need not be a call for a secular philosophical analysis of whether one’s survival was comparatively unfair or not. Instead, it can also be a more existential question about the higher purpose of one’s survival and role in God’s plans. This also helps to explain why it makes sense to ask this question again and again—because it is hard to understand what this higher purpose is. If, however, the “Why me?” question was just an accusatory demand for self-justification, then it would make little sense to ask this again and again—unless it just collapses into a rhetorical way of saying that one’s survival was indeed unfair, in which case, one could just as well keep on repeating to oneself that it is unfair that they survived.

For all those reasons, I remain unconvinced that an appeal to the “Why me?” question could be used as univocal support for the idea that guilt should be modelled as anxiety about warranted resentment and concern about self-justification. I also believe that for this very reason, the “Why me?” question would merit further critical tracking throughout different self-reports because one of the reasons why it is so prevalent may be precisely the fact that it allows for multiple interpretations, and it is an interesting question on its own how often exactly it is that the question is employed in one or the other meaning, or whether there might be any additional interpretations. Addressing those different interpretations is important because, in addition to the phrase carrying a rich potential for meaning, it can also carry a rich potential for misunderstanding.

Furthermore, in addition to the fact that we can imagine cases of guilt without anxiety about resentment and that the “Why me?” question need not lend particularly conclusive support for the idea that survivor guilt in particular should be modelled as concern about resentful calls for justification, there are also more fundamental reasons for why it makes no sense to assimilate guilt with anxiety. Those fundamental reasons would apply even if, in practice, guilt would always co-occur with a relevant type of anxiety because co-occurrence does not exclude distinctness. Most importantly, guilt and anxiety reflect different kinds of core evaluative concerns. Guilt reflects a non-egoistic concern for treating other people well and doing the right thing; anxiety, however, arises from an egoistic concern about receiving a negative response from a relationship partner. To the extent that our understanding of guilt is linked to forgiveness, it is the former kind of emotional response that we expect from others in order to forgive them. We want them to be concerned about not having treated us well because this is bad in itself, not about the prospect of being resented. Their anxiety about having to

face resentment or other negative reactions would only prove that they are concerned about themselves.<sup>89</sup>

Does the fact that we are talking about *warranted* resentment change anything? After all, Velleman does try to improve upon the Freudian account of guilt by introducing a normative element. He does not just talk about vulnerability and anxiety about resentment, he talks about *normative* vulnerability and anxiety about *warranted* resentment. First, it is worth clarifying that the connection with morality does not in and of itself come from the idea that resentment is *warranted* (fitting). It comes from the definition of resentment itself—as an emotion that involves a bitter sense of injustice. The fact that resentment is warranted is only of auxiliary importance—it grants that the other person would not just be falsely accusing us, but that we are indeed morally in the wrong. But, of course, if guilt is a matter of anxiety about warranted resentment, and resentment is to be defined in terms of injustice, then it would still seem that the kind of anxiety that Velleman is talking about involves a reference to a moral concept, and thus the question arises—is this not sufficient to make it a moral emotion?

The answer is still no because the emotion does not reflect a concern about morality for its own sake. An unscrupulous criminal, too, may be anxious about the prospect of getting rightfully caught, but this does not change the fact that they are only concerned with their own well-being. That one does not have anything to say in defence of themselves simply reflects that they cannot appeal to morality in order to shield themselves from negative reactions—they cannot make use of morality for self-defence. But this does not reflect a sincere concern for morality, it is simply a case where an appeal to morality is linked to self-justificatory reasons.

In fact, this takes us back to my discussion (in section 3.1) about how people may be prone to overmoralise the shapes of certain emotions because, for self-justificatory reasons, it is more beneficial to insert a moral claim to the core relational theme of the emotion. For example, it may be beneficial for us to think that our anger involves a sense of moral offence because this would bolster our claim against the one with whom we are angry. For this reason, I called such emotion types pseudomoral—they may be prone to present themselves to us as moral emotions, and may often even involve an explicit thought with a moral content, but at the bottom, they do not reflect a true concern for morality. The dynamics of anxiety about warranted resentment are somewhat similar. Only, in this case, we do not make use of morality in order to bolster our claim. Instead, it is the

---

<sup>89</sup> Somewhat similar ideas have also been expressed in Max Scheler's essay "Repentance and Rebirth", where he reviews the problems of a "fear-theory" of repentance (Scheler, 1987, pp. 105–106). Unfortunately, it is beyond the scope of this dissertation to engage with Scheler's ideas in detail because Scheler speaks in slightly different terms—not guilt but repentance, and is especially concerned with a religious understanding of repentance. As such, he is focused on repentance as a precondition for forgiveness from God, whereas I my aim is to clarify a more secular understanding of guilt, which is more prevalent in the context of interpersonal relationships.

opposite, we are concerned about morality because, in this particular case, it weakens our claim. If we were faced with resentment that is not warranted, we could defend ourselves by saying that resentment is misplaced and nothing unjust has taken place. However, it makes sense to be particularly concerned about warranted resentment because we are especially defenceless in this case—we cannot shield ourselves with an appeal to justice. But morality just figures as a useful instrument in those considerations—as a shield to avert negative responses. Anxiety itself, nevertheless, remains an egoistic emotion that is concerned with the effect that other people’s negative responses might have on one’s own well-being.

To be fair, I do not mean to suggest that anxiety is a totally superficial reaction. Perhaps guilt and anxiety can complement each other. So far, I have emphasised that anxiety about resentment without a sense of guilt that reflects true concern for justice remains somewhat hollow (remember the analogy with an unscrupulous criminal who is worried about rightfully getting caught). But it may also be argued that guilt alone, without any anxiety about resentment (or other negative responses), has the danger of becoming too impersonal. Anxiety recontextualises the matter in terms of more personal goals. It can carry the message that “You are a person of value in *my life*, and I do not want to lose you”.

To come back to Velleman’s dual conception of guilt that I earlier labelled as Janus-faced, it does not seem correct to say that these are indeed the two faces of guilt itself. It would be correct to say that the scenario of having done something wrong (I will model it on the example of traditional guilt because we have not yet resolved all the questions surrounding survivor guilt), elicits two different kinds of emotion—guilt that looks back at what one has done, and anxiety that looks into the future and is concerned about the kinds of social sanctions that one may now have to face. There is indeed a sense in which anxiety about resentment often correlates with the state of guilt but when we talk about the emotion of guilt, we do not just mean any complex of emotions that correlates with the state of guilt, we are more specifically concerned with only part of the experience that reflects a true concern for morality.

I doubt that a Freudian would be satisfied with my analysis, though. As already mentioned, I think it takes a more radical paradigm shift to abandon the anxiety-based account of guilt. If it is of any help, then I suspect that the true disagreement between the Freudian account of guilt as a matter of anxiety about some sort of sanctions and the modern conception of guilt as reflecting a more altruistic concern is ultimately reducible to one’s views about human nature. The Freudian picture, which also has its roots in the works of Friedrich Nietzsche, tends to take a pessimistic view of human nature and posit that people are, first and foremost, guided by selfish and aggressive impulses. If this is the case, then inevitably, guilt must be reducible to those kinds of impulses. Indeed, when we look at Nietzsche’s “On the Genealogy of Morality” (2007) and Freud’s “Civilization and Its Discontents” (1962) we can see that both are concerned with revealing that what may first appear to us as noble sentiments of morality are at bottom just disguised forms of more selfish impulses. Saying that guilt is

ultimately reducible to fear of some negative sanctions is just one way of expressing this more general idea.<sup>90</sup> If it was indeed the case that our idea of guilt as a reflection of a true moral sense is just an illusion, then such folk conception of guilt could be viewed as an example of what Adam Morton (2013) has called an ‘imaginary emotion’—the kind of emotion that, despite the fact that we are used to ascribing it to oneself and the others, does not actually exist. As such, it would be similar to the emotion that Morton calls disembodied, or lustless, passion. The idea of disembodied passion is common in many romantic works of literature, like Goethe’s “The Sorrows of Young Werther”, where a young man perceives themselves as pained by such passion. Self-attribution of disembodied passion involves the thought that one’s love for a (oftentimes married and unreachable) woman is of a completely pure and brotherly nature, unsullied by sensual desire. Morton is sceptical that this is the case and suggests that such characters are just self-deceived about what it is that actually motivates them. The Freudian account of guilt in terms of anxiety suggests the same—even though we may think of ourselves as committed to morality, we are actually driven by more basic impulses.

However, I do not think that guilt, as a matter of moral motivation, is just an illusion. I work within a more optimistic framework, which leaves more room for the possibility of true empathy, altruism, and non-egoistic motivation in general.<sup>91</sup> It is beyond the scope of this dissertation to prove this basic starting point. After all, my main concern lies in the possibility of providing a more inclusive account of guilt that would extend beyond cases of moral wrongdoing. I do not mean to doubt that traditional instances of guilt do not reflect sincere concern for moral wrongdoing, as the Freudian approach would suggest, I simply want to show that reference to moral wrongdoing is not necessary for guilt, and that guilt can be construed as a moral emotion in the absence of moral wrongdoing. However, to the more radical sceptics, let it be said that even if it was indeed the case that we are naïve about the true nature of guilt, I believe that my dissertation will nevertheless have the merit of capturing at least the contemporary *concept* of guilt—that is, how people tend to think of what guilt is about and as such will be relevant to the question of whether our practice of extending guilt to cases without wrongdoing reveals a coherent concept or just reflects a more loose network of perceived family resemblances.

---

<sup>90</sup> There is also another conceptualisation of guilt—the idea that guilt is cruelty turned inwards, aggression against the self. In a sense, at least when we look at the works of Freud (1962), the two views may be seen as complementing each other—guilt is cruelty turned inwards to the extent that it requires a harsh superego who condemns one’s ways, yet from the viewpoint of the ego, guilt involves anxiety about receiving such hard treatment by the superego.

<sup>91</sup> I do not expect this to settle the debate but, for further empirical evidence for the existence of true altruism, see (Warneken & Tomasello, 2009) and (Stich et al., 2010, pp. 157–169).

#### 4.2.2. MacKenzie and Zhao on guilt as a feeling of normative vulnerability

But what if we just emphasised the element of normative vulnerability? This is what Jordan MacKenzie and Michael Zhao (2023) do in their recent revival of the Vellemanian account. The element of normative vulnerability draws attention to the fact that the subject has nothing to say in defence of themselves. As MacKenzie and Zhao understand guilt, guilt is “a feeling of being unable to justify ourselves to others” (p. 2707).

Let me begin by noting that I have no intention to contest the idea that if you feel guilt about a particular matter (be it either wrongdoing or unjust privilege), then you would have to agree that you cannot justify yourself. This is because guilt involves perceiving something as morally problematic, and for something to be morally problematic entails that there is no good moral justification for it. The problem arises from the fact that MacKenzie and Zhao speak about the *feeling* of being unable to justify oneself to others, which implies that the subject is not so much concerned with the morally problematic state of affairs itself but rather with the fact that they cannot justify themselves. Even though MacKenzie and Zhao do not explicitly emphasise the element of anxiety, formulating guilt in terms of normative vulnerability nevertheless does so in an implicit manner. The focus is on the fact that one is defenceless. A sense of one’s defencelessness arises from an egoistic concern about what may now happen to one, about what one would say in defence of oneself when confronted with other people’s negative reactions. The very idea of being defenceless implies the possibility of a looming threat against which one must defend oneself. The very word ‘vulnerability’ refers to the capacity of being wounded—the etymology of ‘vulnerability’ traces back to the Latin word ‘vulnus’, which means ‘a wound’. We talk about a sense of vulnerability when we are concerned about being harmed, so we must already be anticipating a potential threat. As such, the very notions of vulnerability and defencelessness make little sense without the reference to some looming threat, however vaguely it may be imagined.

I imagine that the reason why MacKenzie and Zhao want to emphasise the element of having nothing to say in defence of oneself is that they want to emphasise the social aspect of guilt. This might be promising for interdisciplinary research because it links well with social psychology that puts a great emphasis on how guilt functions in the context of interpersonal relationships. While I am sympathetic to the idea of offering an account of guilt that would be appropriate for not only philosophy but for other sciences as well, I am worried that being overly focused on the social aspect may have its drawbacks. Let me explain. Social psychology is not only concerned with the first-personal experience of guilt. Oftentimes, it looks at guilt from a third-person point of view, emphasising how guilt plays an important role in regulating and keeping good relationships. But the fact that guilt has this larger function of helping to maintain good relationships does not necessarily mean that, from a first-person standpoint, the desire to maintain good relationships must be the root concern of guilt. The reason

why guilt works so well to maintain good relationships is precisely the fact that it does not just reflect the subject's egoistic concern that others be their friends, but instead, it reflects a sincere concern for the well-being of other people. In philosophical and phenomenological inquiry, we are usually more interested in this first-personal experience of guilt, and this is what the talk about the core relational theme of different emotions is supposed to reflect as well. As such, I find it best to keep the core relational theme of guilt attached to non-egoistic concern for justice or other forms of moral badness.

This is linked to how evolution has come to favour altruism. Altruistic people have better relationships with others and, as such, are more likely to both survive and thrive in life. In that sense, being altruistic is usually good for the person. However, this does not mean that altruism itself would have to be concerned with pursuing one's own good. It brings benefits for the person precisely because they are not only looking for their own benefit and are thus more trusted by other people.<sup>92</sup>

By all this, I do not mean to suggest that the social surroundings of guilt should be dismissed. The fact that guilt focuses on the moral is not in opposition to the fact that it is most usually felt in the context of close relationships. Firstly, moral demands are higher in the context of close relationships—we perceive ourselves as owing more to the people with whom we share a long history and a close tie. Since we have more intimate knowledge of our loved ones, we are in an epistemically privileged position to benefit them (Sadler, 2016, pp. 126–127). Also, given their long-lasting and wide-scope nature, close relationships also provide a context where lots of special agreements, promises and commitments are made, both expressly and tacitly (Sadler, 2016, pp. 128–129). Furthermore, even the mere fact that the other has been so committed to us, can be perceived to generate a duty of reciprocity to respond with the same (Baumeister and Wotman 1994, 121–5). So, there are indeed plenty of reasons why a close tie can heighten the demands of morality. Secondly, there is also an important sense in which our attention, and thus also guilt, is guided by our relationships. We might have been unkind to a stranger but, if later in life we will never even meet the stranger, our attention will not remain on this incident, and we are unlikely to keep on feeling guilt about it. Things are different if we have done something morally wrong within a relationship that is an important part of our life: since our attention is on that relationship, we will often be reminded of our guilt. Similar ideas are also reflected in Gibbard's (1990) words: "How fully to engage one's feelings in a situation, then, is a separate question from how to feel if one's feelings are fully engaged. [---] The norms we accept for engagement may let proximity, love, and acquaintance matter" (p. 127).

Even though I will still have more to say about the Vellemanian accounts of guilt, this concludes my critique of those accounts. The central problem with those accounts is that they try to unify the sphere of guilt by appealing to elements related to anxiety. Even though the experiences of guilt are often accompanied

---

<sup>92</sup> See also (Mancini & Gangemi, 2021, p. 4).

by anxiety, we should take care to keep guilt and anxiety conceptually apart. Anxiety reflects an egoistic concern for the sanctions that one may face because of their unjustified position, while guilt focuses on why their position is unjustified in the first place. It may well be that the demise of the Freudian approach to guilt has resulted in the neglect of relationship anxiety as an interesting emotion in its own right, and we would do well to bring it back in order to better illuminate the overall emotional experience that people often have when they feel somehow unjustified. This may also explain the fact why the account offered by MacKenzie and Zhao feels refreshing, because it brings back an important element in the overall emotional experience that one has when they also feel guilt, but this does not mean that we should make anxiety or related elements like that of having a sense of vulnerability into an integral part of guilt itself.

### **4.3. Survivor guilt and comparative injustice**

Yet putting all the previous criticism aside, I still think that the accounts proposed by Velleman (2003) and MacKenzie and Zhao (2023) provide some highly useful insights into the nature of survivor guilt, which provide a wonderful starting point for clarifying what it actually is that unites survivor guilt with the more traditional instances of guilt for moral wrongdoing. So, in what follows, I will focus on what it is that we can learn from those accounts once we get rid of the unnecessary Freudian deadweight, and, building on those insights, I will ultimately move on to defend the idea that survivor guilt is a matter of being a constitutive part of a comparatively unjust state of affairs.

What is left of Velleman's idea of survivor guilt as anxiety about warranted envious resentment, once we purify it from the Freudian idea that guilt is necessarily a form of anxiety? At first glance, it may seem that nothing, but this is not the case. Velleman speaks of guilt through the curved mirror of anxiety, but even though curved, it is a mirror nevertheless. By clarifying the second-personal attitude that is relevant to the scenario which elicits guilt as a first-personal attitude, he does actually tell us a fair bit of what true guilt is about as well. We learn that the second-personal attitude to cases of moral wrongdoing and survivorship is resentment. In the case of moral wrongdoing, it is an angry kind of resentment; in the case of survivorship, it is an envious kind of resentment. What unites both is that, qua forms of resentment, they involve a bitter sense of injustice (Velleman, 2003, pp. 246–247). So, at the deepest level, the unification work is done not by anxiety but by the notion of injustice. This is the most fundamental element that gives rise to both guilt on the subject's part and resentment on the victim's part. So, if we were to construct a more standard (not anxiety-based) account of guilt based on Velleman's insights, then it would have to look something like the following: guilt is an emotion that involves a painful sense of injustice. In case of moral wrongdoing, one would have to perceive their actions as unjust towards others; in case of survivorship, one would have to perceive their survival as unjust towards others. The core concern is not about the prospect of

sanctions or about having nothing to say in defence of oneself, instead survivor guilt reflects a sincere concern for injustice.

I am sceptical whether an appeal to injustice will allow us to accommodate all instances of moral wrongdoing and thus I will have more to say about how exactly we should formulate a general theory of guilt in the next chapter (section 5.1). However, in the context of survivor guilt, an analysis in terms of justice is spot on, and I want to underscore the idea that we should look no further. As we have already seen, attempts to look further have ended up making survivor guilt into something that it is not. In what follows, I am going to cash out the idea that survivor guilt is a matter of injustice in more precise terms, while remaining mindful of not shifting the focus and making it into something else, as so many other authors have done. Everything I say will be such that everyone should agree with if they agree with the basic contention that survivor guilt reflects a concern for the fact that it is unjust that one made it out alive while others who are no less worthy did not. I will try to capture this idea in the most minimalist way possible without introducing any unnecessary elements, and will finally also show how such a minimalist conception lays the foundation for some of the more complex forms that survivor guilt can take.

#### **4.3.1. Comparative injustice as the core of survivor guilt**

First, let us clarify what kind of justice is involved. This is important because once we pin down the relevant sense of injustice, it should be clear that there is no need to include any further elements like the idea that one's privilege must come at the expense of other people. In fact, I suspect that Baumeister's failure with regard to accommodating survivor guilt might have been due to the fact that he did not make fine enough distinctions within the family of justice. One of the reasons that pushed him towards positing that survivor guilt, too, must involve a sense of harm to others—if only indirectly, by benefitting at their expense—was the idea that injustice alone is not sufficient. Let us take a closer look at his line of thought because it will help us clarify what type of justice is not involved.

In principle, an equity theory of guilt could be formulated in terms of calculations by solitary individuals. That is, people might feel guilty simply because they regard their outcomes as being better than they deserve. Research however, has generally shown inequity guilt in interpersonal contexts, in the sense that the person is overrewarded vis-à-vis some salient other person. Solitary overrewards (e.g. receiving extra change from a pay telephone) seem to result in positive moods rather than guilt (see Isen, 1984). (Baumeister et al., 1994, p. 248)

To put Baumeister's point in more philosophical vocabulary, what he is emphasising here is that survivor guilt and inequity guilt, in general, is not just a matter of noncomparative justice—the subject simply receiving more than their due (either in the form of entitlement or desert). Unless there is any harm to others involved, the mere fact of being overrewarded or treated better than one is entitled to does not tend to elicit a strong sense of guilt. Hence also the reference to harm

to others, even though this harm might not be in any way caused by the subject themselves.

But once we switch to the notion of comparative justice, there is no need to add the extra reference to harm to others because this is already included in the idea of justice being comparative. Whereas noncomparative justice concerns what is due to people without reference to others and can thus also be called intrapersonal, comparative justice concerns what is (at least *pro tanto*) due to people with reference to others and is, in this sense, inherently interpersonal.<sup>93</sup> The classic principles of equality take a comparative form because they attribute the greatest importance to consistency and lack of arbitrary inequalities in how the goods and evils are distributed. As put by Feinberg (1974): “Comparative principles all share the form of the Aristotelian paradigm: justice requires that relevantly similar cases be treated similarly and relevantly dissimilar cases be treated dissimilarly in direct proportion to the relevant differences between them” (p. 310). As such, comparative justice is a purely formal matter—it does not give us any substantial standard for how we should treat people but only tells us to treat like cases alike.

There is also an important sense in which principles of comparative justice can combine with principles of noncomparative justice, like that of desert. For example, often, what we are concerned about is that equally deserving people get an equal amount of goods. This is not the same as saying that people should get what they deserve, as it may well be that everyone deserves more or less than they actually get. This is just to say that people with equal desert should receive an equal amount of goods. In this case, it would be appropriate to talk about desert-sensitive fairness or desert-sensitive comparative justice.<sup>94</sup> MacKenzie and Zhao (2023, p. 2718) tie survivor guilt with this type of injustice, emphasising that typically survivors do not perceive themselves as more deserving of survival than those who did not make it out alive. Harcourt (2023) seems to agree as well, as he speaks of survivor guilt in relation to “unequal distribution of outcomes

---

<sup>93</sup> The labels of ‘intrapersonal’ and ‘interpersonal’ have also been used in this context by Kerah Gordon-Solmon (2015, 2017).

<sup>94</sup> I am knowingly departing from Shelly Kagan’s talk about ‘comparative desert’ that is also employed in the treatment of survivor guilt by MacKenzie and Zhao (2023, p. 2718). On the one hand, I find the notion of ‘comparative desert’ idiosyncratic and linguistically inconvenient. For example, it sounds more intuitive to say ‘this distribution of goods is comparatively unjust/unfair’ than it is to say ‘this distribution does not satisfy comparative desert’. But, more importantly, I am also not particularly convinced by Kagan’s project of reducing all matters of justice to desert because it seems to me that in what he calls ‘comparative desert’ the main role is not played by desert but the general principle of comparative justice or fairness, which remains at least as fundamental as desert, and should thus not be pushed out of the picture. On the latter problem, see the critical articles by Kerah Gordon-Solmon (2015, 2017). In any case, Kagan is well aware that his notion of ‘comparative desert’ is not well received by all, and notes that some may prefer to talk in terms of ‘desert-sensitive fairness’ (which is more or less what I have opted for as well). He simply does not think that the label makes much difference. (Kagan, 2003, p. 98, 2014, p. 351, 2017, p. 302)

which is not grounded in desert” (p. 135).<sup>95</sup> While I do agree that tying the comparative sense of injustice with desert is, for the most part, correct, I would leave open the possibility that perhaps someone might perceive matters of comparative injustice in light of other values as well. For example, at least in the Nazi concentration camps, those were not only deserts that were not taken into account in a comparatively fair manner but also rights because it was not the case that those who survived had a stronger right not to be killed. For this reason, I would suggest that the central aspect has nevertheless to do with comparative justice as such, though it may indeed be the case that we usually measure this injustice in proportion to the satisfaction of desert.

Now that we have clarified that survivor guilt is linked to comparative justice, we can move on to explicating the subject’s relationship to comparative injustice. Even if it is agreed that survivor guilt is a matter of perceiving it as unfair that one outlived the others, it is not completely clear how exactly we should explicate the relationship that the subject has with unfairness. This is where our account of survivor guilt must inevitably be partly informed by the more general account of guilt that we have in mind because the relationship between the subject and the comparative injustice must be explicated in a way that allows us to view it as somehow parallel or similar to merely causal cases of guilt and guilt for regular moral wrongdoing. This is also the step at which theorists usually introduce unnecessary further elements in order to grant unity. But as we have already seen, this often leads the focus away from what is central—the injustice itself. So, the challenge is to keep the content of survivor guilt minimal, but explicate it in a way that allows us to view it as belonging to the same category as guilt for merely causal harm and guilt for regular wrongdoing.

My suggestion is that the best way to do this is by saying that, in cases of survivor guilt, the subject perceives themselves as standing in a (partially) constitutive relationship to the relevant type of comparative injustice. This amounts to nothing more than saying that they are an integral part of the constitutive injustice—the very fact that they made it out alive forms part of the comparatively unjust state of affairs that consists in the fact that while they made it out alive, others who were no less worthy than them did not.

Theories of guilt have been so focused on causality, that ways of constitutive contribution have gone totally unrecognised. Let us remember how Taylor (1985) tried to explain guilt for unjust benefits in causal terms: “If I feel guilty about my privileged position in society due to circumstances of 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” (p. 297). She entirely overlooks the

---

<sup>95</sup> To the extent that Harcourt labels the wider genus under which survivor guilt belongs to “guilt at the undeserved” (Harcourt, 2023, p. 135), he could also be interpreted as tying guilt more exclusively with desert only, so that there need not be a sense of noncomparative justice. But given the above-mentioned more specific formulation and the fact that literally all his examples involve an element of comparative injustice, I have chosen to interpret him as talking about desert-sensitive comparative justice.

possibility that, instead of appealing to causal contribution, a much simpler solution would be to say that one's privileged position is a constitutive part of the comparative injustice. While the nature of hedonic well-being is such that we can normally go against it in a causal manner, the nature of comparative justice is such that we can be at odds with it simply because of the position that we have.

### 4.3.2. Comparative injustice as the foundation of more complex forms of guilt

Admittedly, the account that I offer is a rather minimalist account of survivor guilt. But one of its advantages is precisely the fact that it captures the most fundamental layer that is also present in other forms of survivor guilt that do not align with the traditional account of guilt. For example, at-the-expense guilt that was previously discussed, too, assumes that it is unfair that one outlived the others. People do not report at-the-expense guilt in cases where it is comparatively just that one's privilege comes at the expense of other people. For example, imagine that one received a job offer because they were indeed the most qualified candidate. Their position came at the expense of others in the sense that the fact that the job offer was given to her meant that it was not given to anyone else. But there is nothing comparatively unfair about such situation. So, the basic layer of inequity guilt is there in the case of at-the-expense guilt as well, but at-the-expense, guilt adds to it the sense that one's good fortune came at the expense of other people's death.

But, in addition to at-the-expense guilt, there is also a different, even more insidious, turn that inequity guilt can take: in addition to thinking that it is comparatively unfair that one made it out alive, one may also come to think that they do not deserve to live in a noncomparative, more absolute sense of justice. For example, this sense of survivor guilt is captured in the following paragraphs by Edith Eger:

This embarrassment, this feeling of exile, even in my own community, didn't come from without. It came from within. It was the self-imprisoning part of me that believed I didn't deserve to have survived, that I would never be worthy enough to belong. (Eger, 2017, p. 137)

It's my graduation day. I am forty-two years old. I am graduating with a BS in psychology from the University of Texas–El Paso, I am graduating with honors. Yet I can't make myself walk in the ceremony. I am too ashamed. "I should have done this years ago," I tell myself. What I really mean—the subtext of so many of my choices and beliefs—is, "I don't deserve to have survived." I am so obsessed with proving my worth, with earning my place in the world, that I don't need Hitler anymore. I have become my own jailor, telling myself, "No matter what you do, you will never be good enough." (Eger, 2017, p. 163)

In this case, the person can come to think that they do not have a right to life and that it would be, not just comparatively, but non-comparatively and all-in-all

better if they died. This is what may happen if it is forgotten that comparative justice should not be our all-in-all standard. The slave owners, too, might have been comparatively very just towards their slaves, so that they treated them all equally badly, but this does not mean that the slaves would have anyhow deserved such treatment or that the treatment that they received would have been right in a more absolute sense (Feinberg, 1974, p. 312). Similarly, it may indeed be comparatively more fair, if one had not survived the catastrophe, but this does not mean that their survival would not be all-in-all better.

So, on the one hand, according to the principles of comparative justice, it may seem more fair if one had died with the others. But on the other hand, according to the principles of non-comparative justice, it is only just that one has survived. It is a case where the principles of comparative and non-comparative justice come into conflict. Feinberg has aptly captured the confusion that such cases<sup>96</sup> tend to elicit:

The more we ponder cases of the above kind, the more confused we become; for justice and injustice seem alternately to flit in and out of focus like the pictures in an optical illusion, depending upon whether we consider comparative or non-comparative factors. (Feinberg, 1974, p. 314)

Even if one does indeed believe that, in a non-comparative sense, everyone should live, it is only natural that when they think back to their comrades, it is the comparative sense of justice that becomes more salient. As such, it is easy to understand how people can oscillate between these two conflicting ways of looking at the situation, and how matters of justice can seem rather confusing. Perhaps it is also at this point where philosophical analysis of survivor guilt could be of help—in helping to put comparative justice into a wider perspective, where we can see that it cannot be taken as our all-in-all standard of measure. Differentiating inequity guilt from no-right-to-life guilt is an important matter not just conceptually, but also for the purposes of preventing inequity guilt from taking this more insidious form.

---

<sup>96</sup> He says these words about a slightly different situation but one that nevertheless belongs to the same type—a situation where the principles of comparative and noncomparative justice come into conflict. The situation that he discusses arises in the context of the Augustinian theory of salvation, according to which no man deserves to be saved from hell-fire. This means that according to the principles of noncomparative justice, no one deserves to be saved, and has reason for complaint if they are consigned to hell-fire. However, if a selected few are nevertheless saved by Divine Grace, then we can still imagine that those who are consigned to hell-fire will feel that it is comparatively unfair that some lucky few who were no more deserving than them were nevertheless saved. (Feinberg, 1974, pp. 313–314) The situation that he describes is not exactly the survival situation because it is assumed that on non-comparative grounds everyone deserves hellfire. But we can adjust the scenario: everyone deserves to live but God only selects a few to make it out of the concentration camps alive. That those few are saved is itself non-comparatively just because they deserved to live. However, on comparative grounds, it was unjust because the others were not given that chance.

The ease with which people may slip from acknowledging comparative justice to thinking that they have no right to life only goes to show how important it is to acknowledge that while at the very bottom, survivor guilt can be a rational reaction to being a constitutive element in a comparatively unjust state of affairs, this should in no way imply that one's survival is unfair by a non-comparative standard or that one does not have a right to life.

In this section, I have argued that, at its core, survivor guilt is a matter of perceiving oneself as a constitutive part of a comparatively unjust state of affairs, which consists in the fact that while one survived, the others who were no less worthy did not. This is the most fundamental form of survivor guilt because, in the simplest survival scenarios where life is not a zero-sum game, it is the comparative injustice itself that is the sole focus, not the idea that one's survival must have come at the expense of others, nor the idea that one must necessarily think that they have no right to life. However, it is easy to see how those further forms of guilt can build on top of this basic layer, which only goes to prove its essential role.

In this chapter, I hope to have shown that attempts to make sense of the philosophically most challenging form of survivor guilt—inequity guilt—, have not been particularly successful because they have missed the most fundamental layer of survivor guilt which consists of the fact that the subject perceives themselves as a constitutive part of a comparatively unjust state of affairs. This is the most essential aspect of survivor guilt that, on the one hand, lays the foundation for more complicated forms of survivor guilt like at-the-expense guilt and no-right-to-life guilt, and, on the other hand, also gives rise to various additional elements in the survivor's overall emotional experience that, in addition to guilt, may also involve anxiety about envious resentment and a sense of not being able to justify oneself. As it was seen, other accounts have put the focus on those further elements, but this has only led the focus away from the most important idea that, at its very core, survivor guilt involves a sense of being a constitutive part of a comparatively unjust state of affairs.

In a contemporary context, drawing attention to this most fundamental layer is of crucial importance, because, if we want to talk about survivor guilt in a broad sense that also applies to cases where one cancer patient survives while the other does not, or in cases where one is a lone survivor of a plane crash, then we cannot only rely on conceptions of survivor guilt that have been developed on the basis of guilt that was experienced by the Holocaust survivors back when the term was first coined in the 60s. Those cases were highly specific cases where the assumption that one's survival came at the expense of the others was indeed justified because, in a concentration camp, life was indeed often a zero-sum game, and death quotas were a real thing. Those were also highly specific cases because most of the Holocaust victims were Jewish and, as such, shared a unique religious background, which involved a belief in an inherently just world. This explains the cognitive dissonance that is reflected in their experience of the "Why me?" question that arises from the clash between the belief in an inherently just world

and the seeming injustice of one's survival. I am not suggesting that those aspects are exclusive to concentration camps only. Undoubtedly, there is a lot to learn from those cases that is applicable to many other cases as well. But I do want to suggest that these nuances are not always necessary for survivor guilt, especially if guilt is experienced from a more secular perspective and in a situation where life is not perceived as a zero-sum game where one's survival must necessarily come at the expense of others. Also, as we were able to see, some of the nuances may acquire a rather different interpretation. Whereas, in a religious context, the "Why me?" question can reflect existential puzzlement about the higher purpose of one's survival, in a more secular context, it can turn into a rhetorical form of self-accusation. Those are all important differences that we need to be careful about. But my minimalist account of survivor guilt as a matter of being a constitutive part of comparative injustice nevertheless forms the basic layer on which all those different forms of experiences can build on.

Last but not least, an important aspect of my minimalist account is also the fact that I explicate the relation between the subject and comparative injustice as a constitutive one. This may seem obvious, but it has not really been explicated in those terms. Instead, many theorists like Baumeister and Taylor have tried to explain survivor guilt and inequity guilt in reference to causal relations, which I consider fundamentally mistaken. As it will also be seen, explicating the simple truth that we are not dealing with causation but constitution will become of crucial importance when formulating a more general account of guilt that emerges from the treatment of cases for merely causal harm and survivor guilt, which will be the topic of the chapter that now follows.

## 5. THEORIES OF GUILT

In this chapter, I am going to extrapolate a more general theory that emerges from the previous discussions of guilt for harm to others and survivor guilt. I will argue that the kind of guilt that those cases reflect can be captured in terms of feeling bad about having contributed, either causally or constitutively, to a morally bad state of affairs. Since the focus is on the value that is attributed to states of affairs rather than deontic moral wrongfulness, I will call this a value-based account of guilt. In the first part of the chapter, I will unpack what exactly such an account entails and how it applies. Among other things, this will also involve clarifying the ‘contributive’ relationship between the self and the morally bad state of affairs in terms of counterfactual dependence and difference-making. The second part of the chapter is dedicated to situating the value-based account of guilt in the broader landscape of guilt theories and highlighting its advantages. As I will argue, one of its great attractions is the fact that it offers a kind of midway between the traditional account of guilt and the interpersonal one—it combines the outward-looking element of the interpersonal account with the traditional idea that guilt is a moral emotion. I will also argue that, compared to other accounts that take a more value-based direction, mine stands out in that it is not restricted to accommodating only causal forms of contribution but is broad enough to cover a constitutive relation to a morally bad state of affairs, which is the key to understanding survivor guilt. Whereas the second part is focused on the attractions of the value-based account of guilt, the third part zooms in on its limitations. Most importantly, I will explain why the account should not be taken as a theory of all guilt but only reflects the kind of guilt that arises from a value-based moral outlook that is focused on the goodness of states of affairs as a normative dominant, and derives the right from the good. On the one hand, I will expand on which kinds of normative theories such moral outlook best resonates with. On the other hand, I will also illustrate why other normative theories can give rise to a very different kind of guilt, which is not so much oriented towards what happens out there in the world but is instead concerned with the agent themselves being morally stained by their wrongful actions. Last but not least, I will also touch upon the prospects of formulating a unified theory of all guilt.

### 5.1. The value-based account of guilt

The analyses in the previous chapters yielded the following results. Guilt for merely causal harm is a matter of the subject functioning as a part of the causal chain that led to a morally bad state of affairs, which involves undeserved harm for the victim. And survivor guilt is about the subject being an integral, or, in other words, a constitutive, part of a morally bad state of affairs, which lies in the fact that it is comparatively unjust that the subject survived while others who were no less worthy than the subject did not survive. So, now we are finally at a point where we can extrapolate a more general concept of guilt that emerges from those

analyses: guilt as a painful feeling about contributing, either causally or constitutively, to a morally bad state of affairs. In what follows, I will first take a closer look at the account and its applications. In doing so, I will first clarify the three core elements of this account: the reflexive element of the self, the morally bad state of affairs, and the crucial relationship between the two, which I have chosen to capture with the term of art ‘contribution’. Secondly, I will further clarify the substance of the account by considering how its essence can be captured by some metaphorically powerful alternative phrasings, like concern about one’s moral footprint and a sense of having violated the moral order. And, thirdly, I will explain how the account applies to other cases besides those of merely causal harm and survivor guilt—especially the more traditional cases that do involve moral wrongdoing.

### 5.1.1. The elements of the value-based account of guilt

I will begin by unpacking the elements of the value-based account of guilt: the *self* who contributes to the morally bad state of affairs, the relationship of *contribution* between the self and the morally bad state of affairs, and the *morally bad state of affairs* itself, which is the key factor in calling the account value-based rather than norm-based.

First, let us look at the sense of self that is involved in the value-based account of guilt. The sense of self is probably the most tacit element of the account because the subject is more focused on what has happened out there in the world rather than on themselves. But, of course, one cannot just stand in a relationship of mere spectatorship with respect to the morally bad state of affairs—guilt has to involve a reflexive element that refers back to the self as somehow contributing to the bad state of affairs. I do not think that a particularly detailed account of the self can be offered because our guilt feelings can only depend on a subjective sense of self, which may be different for different people. But let us nevertheless map out some options. While for some, this subjective sense of the self, may indeed be the Kantian sense of self as a moral agent who is responsible for their actions, I doubt that most people think of themselves in such narrow terms. As put by George W. Harris (1997): “In terms of what we actually value in ourselves and others [---] we are far more Greek than Christian, despite the long tradition of Christianity and its more recent Kantian progeny. Moreover, our deepest values reflect the fact that we value ourselves as natural organisms, as animals, rather than as gods that transcend nature” (p. 2). While I would not outright deny the possibility that some people may think of themselves in terms of the Kantian sense of the self as a morally responsible agent, I nevertheless think that Harris is right in suggesting that most people’s sense of self is not restricted to their moral agency. In fact, it is also not restricted to one’s agency, which is supposed to be an important element in the so-called ‘agent-regret’ that is usually associated with Bernard Williams’s lorry driver example, where exercising one’s agency in the world by driving a car can, through no moral fault of the subject, result in a fatal accident. As Harris points out, we also care about ourselves as natural organisms,

and I take this to include that we can also identify with our physical body. Thus, one can also feel guilt in cases with no agency whatsoever. Those cases can include miscarriage, where one's body fails to go through with pregnancy. And also cases like those discussed in Talmudic texts, where an immobilised victim is thrown on top of a baby (Juni, 2016, p. 3). If one identifies sufficiently with their physical body, then it will be perfectly natural to feel guilt if their physical body has been used as an instrument of evil. This also applies to the cases where way too many people were forced into wagons during the Holocaust, which ended in some of them getting trampled because there simply was no room in the wagon. Cases like that show how it is possible to feel guilt even for the mere fact that one is a physically extended entity that takes up space. While the fact that we have a body and we take up space might not be part of our *personal* identity that allows us to think of ourselves as unique individuals, this does not necessarily mean that our sense of self would totally exclude those more basic and biological elements.

That being said, it is also important to acknowledge that if one's sense of self is indeed sufficiently narrow and only includes their moral agency, then this may well block them from feeling guilt for merely causal harm and survival, even if they do have a value-based moral outlook and they derive the right from the good. So, for example, I am not claiming that every utilitarian will necessarily feel guilt for merely causal harm and survival (though I would still hold that even if they feel guilt for a narrower range of cases, it would still be value-based guilt that is focused on states of affairs). Given that it is a common practice to extend guilt to a wider range of cases, I have simply assumed that people's sense of self is usually not that narrow.

Secondly, let us look at in what kind of relationship the self stands with respect to the morally bad state of affairs. According to the value-based account of guilt, the subject feels guilt about *contributing* to a morally bad state of affairs. I use contribution as a term of art for the higher-order genus that unites both causal and constitutive relations. There is, of course, a sense in which causal and constitutive relations are different. Causal relations concern diachronic relations between different events, while constitutive relations concern synchronic relations between a part and a whole. Yet it has often been noted that there is nevertheless an important sense in which both relations are similar to each other. According to leading theories in the philosophy of science, this similarity lies in the fact that both relations can be defined in terms of counterfactual dependence: regardless of whether *X* causes *Y*, or *X* constitutes *Y*, it can still be said that *Y* stands in a relation of counterfactual dependence to *X*—if *X* would not have happened, *Y* would not have happened either (Ylikoski, 2013, p. 290).<sup>97</sup> This counterfactual dependence underscores the fact that *X* functions as a difference-maker to *Y*.

---

<sup>97</sup> For an influential counterfactual account of causation, in particular, see also (Woodward, 2003).

I simply use contribution as a more colloquial shorthand to talk about those relations of counterfactual dependence.<sup>98</sup>

So, building on the counterfactual understanding of contribution, to say that the subject contributes to a morally bad state of affairs means that the morally bad state of affairs is counterfactually dependent on the subject. In cases of merely causal harm, for example, car accidents, this means that had the subject not driven that road at that time, the relevant accident would not have happened. In cases of survivor guilt where one's survival is a constitutive part of comparative injustice, this means that had the subject died with the others, there is a sense in which the world would be, at least comparatively, a more just place.

It is important to note that counterfactuals can also help us explain the phenomenon of negative causation, which is especially relevant for guilt because guilt is often felt for omissions. For example, if we adopted a Humean naïve view of causation (Hume, 1974; see also Schaffer, 2004a, p. 203), according to which causes must be physically connected to their effects (like one billiard ball hitting another), we could not explain how one person not paying their debt could be considered as causing another a financial difficulty.<sup>99</sup> But appealing to counterfactuals does enable us to explain those cases: because if the person did pay their debt, the other would not be suffering (at least that severe) financial difficulty. Negative causation is also especially important in cases where one fails to help another—a common source of guilt feelings, especially in catastrophe situations. Had the subject helped the other, the other might still be alive.

Overall, counterfactual dependence underscores the fact that the subject makes a difference to the world, which, in cases of value-based guilt, means that, at least in some sense, they make the world into a morally worse place. Of course, the counterfactual view of causation and contribution is not without its critics. For example, one may wonder about whether the counterfactual account of causation

---

<sup>98</sup> The terms that are used in the theories of explanation in philosophy of science, are not really suitable for the purposes of the current inquiry. I need a verb that would function as a predicate of the subject who feels the emotion, not the predicate of the bad state of affairs. In other words, I need a verb I can use in the following sentence: the subject *f*-s to a morally bad state of affairs (not that the morally bad state of affairs *f*-s, e.g. depends on, the subject). This is important for two reasons. First, for the purposes of continuity I want the verb to have the same relational direction as *to cause* and *to constitute*. Secondly, I find it more natural to start from the explanatory antecedent and to say something about that (the antecedent event causes the final one, the antecedent component constitutes the bigger whole). Thirdly, I also think that this better captures the first-person interpretation of guilt—the subject thinks in terms of how they *f* the bad state of affairs. They do not usually start from looking at the bad state of affairs like a scientist might look at a phenomenon that requires explanation, and then reason backwards, and realise that the state of affairs refers back to the subject. For those reasons, I do not want to talk in terms of the bad state of affairs *depending* counterfactually on the subject. The talk about *making a difference* is better because we can indeed say that the subject makes a difference to the morally bad state of affairs, but I find this phrasing too vague.

<sup>99</sup> See also (Schaffer, 2004, pp. 197–203) for a series of numerous especially convincing examples of how many paradigm cases of causation in general (not only in the sphere of morality) are actually cases of negative causation.

can accommodate cases where the bad result would have happened regardless of the subject. For example, if two people throw a rock at a window, so that one's rock hits the window first, but the other person's rock hits it right after. In this case, it may indeed seem that the throwing of the first rock made little difference to the world because the window would have been broken by the second rock anyway. Yet I do not think that we should be overly worried about those perplexing cases in the current context. We can allow that people who adopt an alternative view of causation, guilt feelings may be construed slightly differently. Neither is it my aim to arbitrate between the more fine-grained nuances between different counterfactual accounts like that of David Lewis (2000; 1971) and James Woodward (2003), because the basic counterfactual shape nevertheless remains the same. In fact, for this very reason, Lewis (2000) has even himself called the finer details of his account mere "extra bells and whistles (p. 182). My modest aim in talking about a counterfactual sense of contribution is simply to capture a conceptual dominant and suggest that, in ordinary circumstances, it seems to be a counterfactual view that people operate with and care about. As such, I am not necessarily interested in developing an accurate empirical account of causation and constitution that scientists should rely on in their work (as, for example, Phil Dow (2000) has attempted). My interest lies in an analytic account of causation and constitution that would simply help us to clarify the relevant folk concept, i.e. what people usually mean when they think of themselves as causing or constituting something. And there is indeed a widespread consensus that counterfactual accounts are especially well placed to explain this and that "the difference-making intuition" (Sartorio, 2005) is indeed an especially pervasive intuition.<sup>100</sup>

One problem that may arise with respect to my appeal to counterfactual relations of dependence like causation and constitution is the problem of remoteness. Namely, it might be suggested that even though it may be common to feel guilt for being the direct cause of a morally bad state of affairs, or with respect to people with whom one was together at the relevant catastrophic event, it does not seem that common to feel guilt about the more remote effects that one may have on the world or with respect to people on the other side of the planet. Indeed, we can be seen as parts of endless causal chains and as constituents of states of affairs that can involve an ever-growing number of other constituents, but the further away we move in space and time, the less common it becomes to feel guilt about one's contribution to the relevant state of affairs. So, the argument might go, my account is too broad and predicts guilt where we usually find none. In response to the problem of remoteness, there are two things I want to note.

First, it seems that, to some extent at least, the lack of guilt for such a broader range of cases can be explained by psychological salience. Most of the time when we go about our daily lives, our attention is not on the remote impact that we have on the world or on very broad-scale states of affairs like those that involve global

---

<sup>100</sup> On the difference between analytic accounts and empirically accurate accounts in the context of causation, see also (Ney, 2014, pp. 325–326).

injustices. Thus, guilt does not happen to arise. However, once one's attention is focused on such issues, guilt may well arise. At least in this day and age, some people do indeed report guilt for the remote effects that they have on the world. It is not uncommon to hear worries about the wide-scale environmental effects of our daily consumption habits, and even of having children. Similarly, given that information technology can make us more acutely aware of people on the other side of the planet, who live in poor conditions and suffer from malnutrition, one can also feel guilt for being unfairly privileged. In fact, my account of guilt can be seen as an especially timely development because it does indeed reflect the type of guilt that has become more and more common these days.

But, secondly, I do agree that there is also an important truth in the idea that people are more prone to feel guilt about the more immediate effects that they have on the world, and with respect to people with whom they have come into closer contact. However, I do not think that any of this poses a problem to my account of guilt. To the contrary, given the important role that the counterfactual account of contribution gives to difference-making, it can be easily explained why it is less common to report guilt for very remote effects that one has on the world, and with respect to very large-scale injustices. It is because these are the cases where the difference that we alone make tends to become increasingly smaller. The further away something stands from us in the causal chain of events, the more there will usually also be other causes that have played a role, so our own contribution in the sense of the overall difference that we alone make tends to get increasingly smaller. For example, it may indeed be true that one has contributed to global warming, but since global warming is a large-scale remote effect, it is only a minuscule percentage of the damage that can be attributed to a single individual. Similarly, if one thinks of oneself as a party in the large-scale injustice that lies in the fact that people in the first world countries enjoy a relatively comfortable and wealthy lifestyle, whereas those in the third world suffer from malnutrition and various diseases, one's own relative privilege is just a very small part of the injustice. There is indeed a sense in which their not enjoying their privilege would make a slight difference to the comparative justice, but it would do so only to a minuscule extent. Whereas if we are talking about a case where the reference group is small, for example, one cancer patient outliving the other and recovering, then the individual makes a proportionally greater difference to the overall comparative injustice in question.

None of this is to deny that people can feel a strong sense of guilt for remote effects and large-scale injustices as well, because the overall amount of guilt that one feels may be amplified by collective guilt. But this is a separate matter, and I am here only comparing the amounts of individual guilt. And when I say that individual guilt feelings tend to be smaller in case of remote effects and large-scale injustices, I am only touching upon a general tendency which need not apply in all cases. For example, remember the man who felt guilt for outdoing Hitler at the school entrance exams because they thought that this might have played an important role in Hitler's hate for Jews and thus also the occurrence of the Holocaust. If this was indeed true, then the subject would be a remote cause, but

nevertheless a potentially very significant cause that might have made a big difference. Similarly, if one is not just an ordinary person living in the first world but belongs to the very small group of billionaires who control most of the world's wealth, then the difference that they make is much bigger. So, ultimately, it is the overall difference that matters, not the immediateness of the effects or the closeness of the reference group. So, the counterfactual idea of contribution is not just a yes-or-no question of whether one contributes to a morally bad state of affairs or not, but also, and even more importantly, a matter of how much difference one makes.

This concludes my positive account of contribution, but before I move on to the morally bad state of affairs to which the subject contributes, I would like to address a question that might arise in relation to the fact that I define contribution in terms of counterfactuals. Given that there is another relatively similar emotion, regret, that has been labelled as “a counterfactual emotion” (Kahneman & Tversky, 1982, p. 206; see also Landman, 1987, p. 142) or even as “the prototypical counterfactual emotion” (Zeelenberg & Van Dijk, 2007, p. 148; see also D. Mandel, 2003), it is worth clarifying whether guilt has the same relationship to counterfactuals as regret, and if yes, then in what sense, if any, is guilt distinct from regret? Let us start with the reason why regret is called a counterfactual emotion. It is because regret depends on counterfactual thought—in order to feel regret, the subject has to think in counterfactual terms and compare the current situation with what might have been (Zeelenberg & Van Dijk, 2007, p. 148; see also Roese, 2001). As such, regret has been considered as a cognitive emotion—it is not just a feeling but an emotion that involves counterfactual thought. It is certainly true that guilt may sometimes be associated with counterfactual thought as well. People who feel guilt, especially in causal cases, may well be burdened with what-if questions, wondering how things could have gone differently had they done something differently (e.g. chosen a different route when driving back home, left a few minutes earlier, and so on). But the value-based account of guilt does not suggest that the subject must necessarily think in counterfactual terms. On the one hand, given that I remain neutral with respect to the theory of emotion, the subject need not think at all—the emotion can be just a feeling without propositional content. On the other hand, even if the subject does think of themselves as contributing to a morally bad state of affairs, this does not mean that they must explicitly engage in counterfactual thought (or imagination)—they can just think of themselves as having caused something bad, or as being part of a comparatively unjust state of affairs. After all, counterfactuals only figure in the definitions of contribution, but in our stream of thought, we rarely define each individual word that goes through our mind. So, counterfactual thought is not an essential part of guilt, and to the extent that it is an essential part of regret, guilt does not always go hand-in-hand with regret either. Yet the value-based account of guilt can easily explain why guilt can often be combined with counterfactual thought—because counterfactuals highlight what it means to contribute to a bad state of affairs, they highlight the fact that the subject makes a difference, and the

difference becomes especially visible when we compare reality with counterfactual scenarios.

Now, finally, we can move on to the most essential element of the value-based account of guilt—the fact that the subject construes oneself as contributing to a morally bad state of affairs.

Overall, normative concepts can be divided into two broad types: axiological or evaluative concepts (from the Greek *axios* and the Latin *valores*, which both mean ‘that which has worth’) like the good and the bad, and deontic concepts (Greek *deon* ‘that which is binding’) like right and wrong (see also Tappolet, 2013, p. 1). Axiological concepts like goodness can be attributed to both states of affairs and character traits. By labelling my account value-based, I intend to explicitly restrict my focus to the kind of goodness that is attributed to states of affairs,<sup>101</sup> not the goodness of character, which might alternatively also be referred to in terms of virtue. So, unless specified otherwise, it is the goodness of states of affairs that I am concerned with. The value-based account of guilt focuses on the idea that the object of guilt is a morally bad state of affairs, and as such, the account stands in contrast to the traditional account of guilt that views the object of guilt as morally wrong and thus operates within the domain of the deontic.<sup>102</sup>

It is also crucial for value-based guilt that moral badness of states of affairs must function as a normative dominant within the subject’s moral outlook, by which I mean that, for the subject, this is the most salient element that functions as a source of normativity. It does not matter if the person is morally responsible for contributing to the bad state of affairs and has thereby engaged in moral wrongdoing. Even if they are, the wrong does not function as a normative dominant but would only derive its negative value from the badness of the relevant state of affairs.

Admittedly, I express myself in somewhat technical philosophical terms when I speak of valuable states of affairs as a normative dominant, or the good being prior to the right. This might make one question whether such a theoretical view is indeed fit to play a role in the kind of wider folk talk about guilt that my account seeks to explain. I believe it is. Of course, I am not suggesting that people should themselves be able to explicate their moral views as value-based. It is precisely

---

<sup>101</sup> At earlier times, some authors like Vallentyne (1987), pace Stocker (1973), did indeed use the word ‘axiological’ precisely to capture only this more narrow focus on the goodness of states of affairs. In this more narrow sense of the word, my account could also be called an ‘axiological’ account of guilt. But given that virtue ethics has made a strong comeback, and goodness is also often attributed to character traits, I have tried to avoid any potential misunderstandings here and steered away from calling the account ‘axiological’.

<sup>102</sup> I have chosen to speak in more neutral vocabulary of moral badness but, to the extent that people talk about natural evil (for example, in cases where a whole settlement of relatively decent people suffer from a natural catastrophe), the concept of moral badness could also be captured by the concept of evil. In colloquial language, both moral badness and evil can be used as antonyms of moral goodness. And in a more theological context, the problem of evil is certainly related to not just manmade evil but to natural evil as well.

because I find it overly pretentious to attribute to people something as explicit and as coherent as a moral theory that I have chosen to speak in terms of a moral outlook. This also leaves open the possibility that moral perception is not just a matter of thinking but that people can also be attuned to morality via non-reflective means such as habit and emotion. Overall, the talk about the goodness of states of affairs as a normative dominant means that one's moral perception is, for the most part, attuned to and oriented towards what happens out there in the world as the most salient feature of the situation when making moral evaluations. Since what is at stake is not just a technicality of moral theorisation but one's basic pattern of attention, I am optimistic that the philosophical talk about the goodness of states of affairs as a normative dominant is sufficiently broad to translate into a psychologically relevant difference.

Next, I would like to clarify why it is that I talk about *moral* goodness/badness in particular. While I am not the only one who ties guilt with *moral* badness—the idea is shared by Zhao (2020, p. 294) as well—a question that might arise with regard to whether it is really necessary to add that guilt concerns *moral* badness rather than just badness as such. When philosophers speak about how we might derive the right from the good, no one seems to add that the right is derived from the morally good, so talking about the morally good may sound unnecessary, and may also be perceived as at odds with the attempt to naturalise morality by reducing moral duties to some good that can be explained in natural terms (e.g. hedonic pleasure). Furthermore, for someone who thinks of morality in Bernard Williams's sense, as restricted to the sphere of moral obligation and the voluntary (B. Williams, 2006, p. 175), it may sound like an outright contradiction in terms to talk about cases where one has, through no act of their own, contributed to a 'morally' bad state of affairs. It might be said that there is nothing moral about states of affairs per se; it is only with regard to voluntary action that we can speak of morality to begin with. So, let me be clear that I am not using the term 'moral' in such a narrowly deontic sense.

Instead, since my aim has been to provide an alternative to the standard view according to which guilt is linked to moral norm violations (which I have viewed as conceptually tied to moral wrongness), I have opted to talk about 'moral' badness simply for the sake of keeping the parallel with 'moral' norm violations. As I explained in section 2.1.2, in the standard account of guilt, the reference to moral norms is best explained by formal descriptive definitions of moral norms, according to which only those norms that the subject perceives as, roughly speaking, especially important, authority-dependent and universalisable (at least in the weak sense of not minding if other adopt the same norms). In this context, 'moral' does not so much refer to the link with voluntary action (norm violations in general can be said to presume that norms are flouted voluntarily), but instead functions to contrast moral norms with other types of norms—aesthetic, epistemic, prudential, etc. In the same vein, we can also speak about moral goodness and moral values—as referring to only those values that the subject perceives as especially important and universal. Here, too, the qualifier 'moral' helps to contrast moral values with other types of values. In the discussions about how the right

can be derived from the good, which already take place within a context of moral philosophy, it may seem unnecessary to add the qualifier that emphasises that we are talking about ‘moral’ goodness. But once we zoom out from those contexts, it is nevertheless important to make a difference because we can also consider beauty as a good, yet aesthetic values are not usually perceived as placing demands on us in a way that moral ones do<sup>103</sup> (unless, of course, we take a moralised view of them, which is compatible with formal notions of morality that do not place material restrictions on the content of moral matters). Even though it is not impossible that someone might report guilt feelings for painting an ugly picture, it is not all that common, and if someone did report guilt for painting an ugly picture, we would be likely to expect some explanation with regards to why exactly it is that they feel guilt about it (for example, because they fail their art instructor, or because they take it as their perfectionist duty to paint beautiful pictures). Building on the idea that we call ‘moral’ only those norms and values that are perceived to place especially strong demands on us, my suggestion is that we can talk about the morally good because it is precisely this type of good that we perceive as grounding the deontic duties that morality in a more traditional sense is concerned with. Beauty does not (usually) place as serious demands on us and ground our deontic duties as values related to moral goodness do—values such as the well-being of other people, autonomy, justice and so on. I suspect that it is the same intuitive link that also explains the practice of talking about natural evil. Evil seems like a synonym for moral badness, and cases of natural evil like catastrophes seem to be exactly the kinds of cases that, if they were results of knowing and wilful human actions, we would consider the people who bring about such catastrophes as having done something morally wrong.

In any case, as argued at length in section 2.2.2, I do not think that the main problem with the standard account is the fact that it makes reference to *moral* rather than just any kind of norm violations. It is not this aspect of the standard account that this dissertation has sought to problematise. The main focus has been on shifting the attention from norms and wrongness to values and badness. So, linking guilt with ‘moral’ goodness is not intended as a particularly new or intriguing statement on its own. I have just relied on the idea that there is no need to fix something that is not broken to begin with.

Even though moral badness itself is a normative primitive that is not reducible to something else, when we look at specific cases of guilt, there is still a lot that can be said about which anti-values, in particular, are perceived as instantiating the morally bad. As shown in my previous analysis, cases of merely causal harm are usually linked with pain as a base-level disvalue, and noncomparative injustice, usually in the form of negative desert, as a higher-level disvalue that gives further negative evaluation to the fact that such a pain is suffered by a good person. Survivor guilt, however, involves a reference to comparative injustice—the idea that equally worthy people received an unequal fate. The connections

---

<sup>103</sup> See also (Steinbock, 2014, pp. 110–115).

with both noncomparative and comparative forms of justice are especially interesting because they help to explain why people often describe the morally bad state of affairs as “not right” (Feinberg, 1974, p. 309). It is not because one would have violated an agent-based deontic duty, but because what happened is seen as unfair towards the victim. So, we are dealing with moral badness in a special sense—it is not just that the end result is undesirable, as it is usually thought of in case of values in general, it is the fact that the end result is perceived as “not right”, as not how things “ought to be”.

The fact that both guilt for merely causal harm and survivor guilt involve a sense of injustice may raise the question of whether all guilt could be viewed as a matter of justice. Remember that, at least in part, Velleman’s idea of guilt as anxiety about warranted resentment (discussed in 4.2, and used as a starting point in 4.3), where resentment was understood to involve a bitter sense of injustice, hinted in this direction as well.<sup>104</sup> If it was indeed the case, then my original formulation of the value-based account of guilt in the more general terms of moral badness would be only necessary but not sufficient to capture the unique evaluative profile of guilt—because we would be dealing with not just any kind of disvalue that might instantiate moral badness but injustice in particular. But does reference to justice really provide us a superior way to unify the sphere of guilt? Even though there are some influential authors throughout history who have indeed tried to reduce morality into a sole matter of justice (Aristotle, 1991a; Kohlberg, 1981) and guilt itself, too, has been associated with a general sense of justice (Deigh, 1982),<sup>105</sup> there is a rather wide contemporary consensus that this is not a particularly fruitful way to talk about morality (Kristjánsson, 2006, p. 42). First, a justice-based account of morality and guilt contains an important loophole. As already noted by John Stuart Mill, we should not “merge all morality in justice” because generosity and beneficence are not matters of giving people their due (Mill, 1998, p. 95). Imagine a utilitarian who thinks they have a duty to give to charity but, having failed to do so, feels guilt. It is not because they violated the demands of justice by that; it is because they failed to promote other people’s

---

<sup>104</sup> But let us not forget, that Velleman himself ends up defending a disjunctive account, which admits two origins of guilt—one that is linked to angry/envious resentment (this is the part that is supposed to cover moral wrongdoing and survivor guilt), and the other that is linked to loss of trust (this is the part that is supposed to cover moral wrongdoing and self-disciplinary guilt). So, strictly speaking, he does not necessarily have to insist that all cases of moral wrongdoing involve anxiety about resentment (and are thus linked to injustice), but could opt for a family resemblance account and suggest that paradigmatic examples of guilt involve both anxiety about resentment and anxiety about loss of love; and less paradigmatic examples where one of those elements is missing must at least involve the other element. In this case, instances of guilt that do not involve anxiety about resentment (a sense of injustice) could be explained in terms of anxiety about loss of love.

<sup>105</sup> John Deigh, in particular, is only interested in justice in a deontic sense though, not telic.

well-being.<sup>106</sup> Another rationale for keeping beneficence and justice, at least in the form of desert, apart can also be found from David Ross, who, after stressing the importance of desert, asks whether we are fundamentally concerned with pleasure and pain at all (Ross, 2002, pp. 135–136). To the extent that people do not want bad guys to enjoy great pleasures, but might prefer them to suffer pain and hardship instead, it may seem that what people really care about is not just the fact that there would be the maximum amount of pleasure in the world, and the least suffering—but the fact that people get what they deserve. So, the thought goes that it is not pleasure and pain as such that are respectively good and bad, but only deserved pleasure is good, and only undeserved pain is bad. In the opposite cases, the values of pleasure and pain are reversed—undeserved pleasure is bad and deserved pain is good. But even if this line of thought is correct, Ross suggests that there is still a rationale for counting pleasure as a *prima facie*<sup>107</sup> good (and pain a *prima facie* evil) because in cases where the notion of desert does not apply, for example when speaking of animals, we would still consider pleasure as good and pain as bad (Ross, 2002, p. 137). This line of thought can be further supported by Thomas Hurka’s recursive conception of moral goodness, where pleasure can be considered as a base-level good, and desert as a higher-level good that derives its value from the combination of good things happening to good people (and bad things happening to bad people).<sup>108</sup> What we can learn from both Mill’s suggestion that beneficence is not a matter of giving people their due, and Ross’s idea that there is room for valuing pleasure apart from considerations of desert, is that the moral domain can become apart from the domain of justice, and thus it would be overly restrictive to reduce matters of morality and guilt to those of justice alone.

---

<sup>106</sup> I am not saying that it is impossible to construe the situation in terms of justice (perhaps one might suggest that giving for charity is an important part of making world into a comparatively more just place). I am saying that for a true hedonistic utilitarian like Mill, who is first and foremost concerned with pleasure and pain, justice is not the ultimate goal, it is valuable only to the extent that it helps people live more pleasurable lives.

<sup>107</sup> In order to avoid controversy by choosing a different term, I have stuck to Ross’s original term *prima facie*, but, admittedly, it is often noted that his notion of *prima facie* is misleading (Stratton-Lake, 2002, p. xxxiii).

<sup>108</sup> There is an interesting difference between Ross’s and Hurka’s accounts, though. Ross would not say that there is anything bad about an evil person suffering a deserved pain (Ross, 2002, p. 136), and within the context of axiological guilt, this would mean that there is no reason to feel guilt. However, Hurka would count the goods and evils differently—he would say that there is nevertheless a *pro tanto* sense in which the suffering of an evil person is a bad thing, just qua suffering, but this is outbalanced by the fact that the suffering is good in virtue of being deserved because of their vicious nature. So overall, a bad person suffering a deserved pain is good, but this does not erase the fact that his suffering is nevertheless *pro tanto* evil, and that, in addition to being glad about the victory of desert, there is also a reason to be sad about the suffering (Hurka, 2000, pp. 194–195). So, in the context of value-based guilt, there might still be reason to feel guilt because one has nevertheless brought about something that is *pro tanto* bad. My account leaves it open whether we should count the bad in Ross’s way or Hurka’s. It all depends on whether the person considers the suffering as *pro tanto* bad or not.

There is also another reason why we might prefer not to speak in terms of injustice when talking about moral wrongdoing in general. This is not in and of itself a knock-down argument (like the previous one) but a practical, though nonetheless important, concern. Namely, there is a sense in which saying that moral wrongdoing, qua consisting in violation of rights, is a matter of injustice does not really add much of substance. Indeed, when we make injustice a general term for all (or most part of) moral wrongdoing, its meaning becomes so diluted<sup>109</sup> that it hardly gives us any clue with regard to what are the more substantial moral values that are at stake in any particular instance of wrongdoing—for example, are we dealing with violations of autonomy, or infliction of pain? This is also what Wojciech Sadurski (1985) must ultimately have in mind when he notes that when we talk about cruelty to animals, obscene language, treason and dishonesty “condemnation in terms of ‘injustice’ hardly comes to one’s mind” (p. 12). So, it is also for this reason that we might sometimes just want to reserve the word justice to matters of comparative justice—because this is a more substantial matter that gives us a clearer clue about what it is that we are talking about (that people with equal merit receive an equal share of the goods). So, there are indeed many reasons for talking about justice as just one moral consideration among the many, and for all those reasons, I do not think that we should completely change the paradigm and translate all our talk of morality into the terms of justice. Instead, we can just view justice as one value among the other values, like pleasure and autonomy, and guilt for moral wrongdoing can also consist in going against those other values.

This completes my formal analysis of the value-based account of guilt that identifies the core relational theme of guilt as a matter of oneself contributing to a morally bad state of affairs. When I state my account in terms of the subject perceiving oneself as contributing to a morally bad state of affairs, I do not mean to suggest that those are necessarily the words that the subject themselves would use to describe what their guilt is about—especially if they are unfamiliar with the theoretical discourse in normative ethics. It is the idea that is important—the idea that is conveyed once we explain what it means for the self to contribute to a morally bad state of affairs. The reason why I speak in terms of the morally bad and causal and constitutive forms of contribution is because those are some of the most fundamental terms in moral philosophy (‘bad’) and philosophy in general (‘causation’ and ‘constitution’). The fact that all those terms have a long and well-established history in philosophical thought, that they are tried-and-true terms, is useful because they help us capture the relevant sense of guilt in an especially clear manner. However, in the next section, I want to complement my perhaps

---

<sup>109</sup> Wojciech Sadurski (1985) makes a similar point about Aristotle’s notion of general justice: “this broader sense of justice (as a ‘complete’ virtue) is so all-encompassing that it serves no useful function at all. If all good acts, rules or societies are ‘just’ in this sense, but some are also ‘just’ in the narrower sense, we might as well restrict the term to its latter meaning and so have the advantage of comparing and contrasting some virtues of those acts, rules or societies with other virtues” (p. 12).

somewhat technical and dry account of guilt with some more colourful ways of expressing the same basic idea.

### 5.1.2. Alternative phrasings: moral footprint and violating the moral order

This section is dedicated to the more metaphorical ways of capturing the idea that the subject contributes to a morally bad state of affairs. In particular, I will focus on the following phrases: moral footprint and violating the moral order. On the one hand, I hope that touching on those alternative formulations will resonate better with those who are perhaps not that familiar with the jargon of normative ethics. Secondly, I am also optimistic that an appeal to metaphors will help to further illuminate what this value-based sense of guilt is ultimately about—even for those who might think that the original formulation will suffice.

Let us start with the idea of moral footprint, which is relevant in cases of causal harm and captures the idea that the subject has caused something morally bad. In this day and age, when environmental concerns are at their peak, it is common to talk about our ‘ecological footprint’, which reflects the impact that people’s way of life has on the environment. Similarly, we can adopt the metaphor of a footprint to talk about our moral impact on the world. The metaphor of a footprint marks a causal way of making a difference in the world. Leaving a footprint means that we step on the ground, creating an indentation in the soil. Our moral footprint is the sum of the goodness and badness of the states of affairs that we have causally brought about, no matter how willingly or unwillingly.

When we look at the cases where guilt is felt for causal harm, it is evident that the subjects are highly concerned about having a positive moral impact on the world. Let us return to the case of Patricia, who thought that the fact that she was not morally at fault did not really explain her guilt because it was nevertheless her who killed the motorcyclist. She seems highly concerned with her moral footprint. This is evident from how she has lived her whole life. She is reported saying: “I spent my whole life volunteering—for animal shelters, for Make-A-Wish [---] This just negates everything good I’ve ever done.” (Gregory, 2017) In fact, it seems that since she has dedicated her whole life to making the world a better place, leaving a positive moral footprint is one of the backbones of her personal identity, an identity-conferring commitment, to use a widespread phrase from integrity literature (Kekes, 1983; see also McFall, 1987, p. 13). As such, it must be an especially hard blow to be involved in such a tragic accident. As she described it, killing another person seemed to undermine all her efforts to make the world into a better place. Or consider a post from the Hyacinth Fellowship web page, written by Gregg who had caused his high school sweetheart to die in a fatal car accident: “Eventually I chose a career part as an expert thought leader, author and consultant on respect that was designed—at least in my mind—to make a positive difference in the world and pay back what I took from the universe. [---] I realized that there is more I could do—specifically making a positive,

lasting difference in the world the way my girlfriend would have IF she had lived”.<sup>110</sup> Since people who feel value-based guilt are concerned about the impact they have had on the world, this also points towards a constructive action tendency—even if they cannot undo the harm, they can nevertheless contribute to a better world. Here, it is especially interesting to note that not only is Gregg concerned with the more immediate consequence of causing a death but also the more long-term consequences that followed. Thus, he thinks about the positive impact that he believes his girlfriend would have had on the world, had she not died in the accident, and seeks to compensate for that as well. From the same Hyacinth Fellowship support page, we can also read of Susie, who had run over a motorcyclist, who reports: “I’m trying to focus on making my life worthwhile, doing my best for others. I plan on donating blood for the first time”. All those stories tell us about the concern that people who feel value-based guilt have about the impact that they have had on the world. They are concerned with their past moral footprint, and they are also concerned with their eventual moral footprint on the world. Even though they cannot undo the harm that they have caused, they can nevertheless pursue a constructive outlet for their feelings of guilt and find relief in trying to make a positive difference to the world.

Secondly, I want to elaborate on the idea of violating the moral order, which, as I will soon show, has the potential to accommodate not just cases of *causing* a morally bad state of affairs but also cases of standing in a *constitutive* relationship to morally bad states of affairs. But let us start from where the phrase comes from. My talk about violating the moral order is in great part inspired by how guilt is described in the works of Janice Lindsay-Hartz (Lindsay-Hartz, 1984; Lindsay-Hartz et al., 1995). For philosophers, or at least those philosophers interested in conceptual clarification, those studies are perhaps some of the most interesting ones out there about guilt because Lindsay-Hartz uses a phenomenological method called conceptual encounter, which involves comparing the investigator’s conceptualisation of a phenomenon with the participants’ own concrete experiences of the phenomenon. As a result, they arrive at emotion descriptions that are part of an interactive process between the investigator and participants. (Lindsay-Hartz, 1984, p. 690; Lindsay-Hartz et al., 1995, p. 275) Furthermore, their goal was an ambitious one—the aim was not just to describe the different varieties of guilt based on family resemblance but “to develop abstract descriptions of guilt [---] that would describe *each and every example* of guilt” (Lindsay-Hartz et al., 1995, p. 276). What is interesting about the results of her study is that she opted to talk about guilt in terms of “a violation of the moral order” (Lindsay-Hartz, 1984, p. 698; Lindsay-Hartz et al., 1995, p. 278), sometimes also speaking in terms of how the happening of bad things “upsets” the moral order (Lindsay-Hartz et al., 1995, p. 698). Admittedly, the studies of Lindsay-Hartz were not meant to accommodate cases without wrongdoing, but it is nevertheless striking how the phrasings that she arrives at can capture a broader sense of guilt than the

---

<sup>110</sup> <https://hyacinthfellowship.org/blog/join-the-conversation-comments-from-our-community/>, visited March 10, 2025.

one that would just involve a violation of the deontic order (which would basically just amount to violating deontic moral norms). Especially to the extent that guilt usually involves a sense of injustice, the talk about violating the moral order can be applied to value-based guilt as well. Someone who has killed another person in a car crash might well construe herself as violating the moral order of the world in the sense that it is still unfair toward the victim to have lost their life. But even more importantly, reference to violating the moral order also captures the sense in which someone feeling survivor guilt might describe their experience: they feel that the very fact that they came out of the catastrophe alive and continue to live on violates the moral order because equally worthy people ought not to receive so drastically different fates. The very fact of their continued existence is a standing violation of the principle of comparative justice. What is meant by the *moral order* in cases without moral wrongdoing is one's sense of how the world should be. This takes us back to Feinberg's idea, which, even though proposed in the context of noncomparative justice, applies equally well to the telic understanding of noncomparative justice. He draws our attention to the fact that it is possible to speak about states of affairs as "not right": saying that "It is not right that such and such should be the case" involves "a stronger and sharper complaint than simply stating that it is not a good thing that such and such is the case" (Feinberg, 1974, p. 309). Stating that something is not right can involve a reference to the idea that it is not in accordance with the principles of justice. It is in that sense that our good fortune can sometimes violate the moral order.

Overall, both alternative wordings—the subject leaving a negative moral footprint; and the subject constitutively violating or upsetting the moral order—highlight the very same idea that they contribute to the morally bad state of affairs in the sense that they make a difference to the world, and without their relevant contribution, the state of affairs would be at least somewhat less bad.

### 5.1.3. Further applications

This dissertation as a whole has been focused on guilt for merely causal harm (chapter 3) and unjust benefits (chapter 4). In this section, I will further apply the value-based construal of guilt to two further types of cases: cases of traditional moral wrongdoing and a special form of spectator guilt. As I will argue, the first builds on a causal relation to a morally bad state of affairs, whereas the second builds on a constitutive relation.

So far, I have primarily focused on how the value-based account of guilt allows us to accommodate experiences of guilt in the absence of perceived wrongdoing. Now, let me explain how it can also cover cases of traditional wrongdoing. When I talk about cases of traditional wrongdoing here, I do not mean each and every experience of traditional guilt that involves a sense of moral wrongdoing. As I will explain in greater length in 5.3.2, the account does not align well with certain forms of deontology and thus will not function as a theory of all guilt. But here the focus is on the extent to which the account does have the means cover

guilt for moral wrongdoing—as felt by people with a value-based moral outlook, that is, for those who consider the good as prior to the right and are thus especially attuned to the goodness/badness of states of affairs. And this is nevertheless important to show because, overall, my aim has been to argue that the kind of folk discourse that applies the word ‘guilt’ more liberally, to cases without moral wrongdoing, does capture at least a subjectively coherent guilt concept. And if the concept is to be coherent, then it will have to capture the experience of both guilt for moral wrongdoing and guilt without moral wrongdoing. It will have to show how the two can be construed as species of a genus, so as to merit the same label. Otherwise, we are not dealing with a unified concept but just a loose set of family resemblances.

The key to understanding guilt for moral wrongdoing is its link with *causing* a bad state of affairs. This may seem simple enough in the case of violating consequentialist moral norms—one’s actions are seen to cause harm and suffering. But the idea that traditional violations can be construed in causal terms goes beyond the idea that it is just in virtue of the consequences of our actions that there is a causal relation between the subject and the morally bad. It is important to note that our idea of moral agency is usually built on the causal theory of action,<sup>111</sup> according to which actions are events that are caused by our mental states like intentions—a view that is “by far the most widely accepted view in the contemporary philosophy of mind and action” (Schlosser, 2019, p. 22) and which seems to align with our folk understanding on action as well (for the conceptualisation of emotions, it is ultimately the latter that we are concerned with). So, in addition to the fact that we view our actions as causing further consequences, we also see ourselves as causing our actions by our mental states. This enables us to extend the account of guilt to at least some deontological violations as well. If we rely on a value-based construal of deontology, according to which norms are derived from the good, where the good can also be justice or autonomy, we can say that certain actions are intrinsically wrong because the event of the action itself is a morally bad state of affairs. If I lie to another person, then this is inherently bad because, regardless of any further consequences, the very act of lying is a violation of the victim’s autonomy. If I distribute goods in an unfair manner, then this, too, regardless of any further consequences like whether it makes the recipients sad or not, is inherently bad because the unfair distribution itself is morally bad. In all those cases, the subject can be said to causally contribute to a morally bad state of affairs because the subject themselves is seen as a cause of their action, which in turn is seen as morally bad.

Perhaps it may seem weird to make traditional instances of guilt that make use of moral culpability a special case of guilt. It might be suggested that I have started from the wrong end: I have constructed an account of guilt based on more unusual cases of merely causal harm and survivor guilt, and only now try to

---

<sup>111</sup> For influential examples of the proponents of the causal theory of action, see (Davidson, 1980a, 1980b; Goldman, 1970; Brand, 1984; Bratman, 1987; Dretske, 1988; Bishop, 1989; Mele, 1992, 2003; Enç, 2003).

provide an *ad hoc* solution to accommodate the more traditional instances that do involve moral wrongdoing. It may seem that by doing so, I make the traditional guilt a more borderline kind of special case. But such worries are unfounded. The value-based account of guilt has all the resources to explain why guilt for moral wrongdoing is nevertheless the most paradigmatic form of guilt. This is so because agency is an especially important and central aspect of our sense of self. While people often do identify with their physical body as well, we do not usually focus on ourselves as merely physical objects that can crash with other physical objects—just like billiard balls on the pool table. In ordinary circumstances, we think of ourselves as agents who lead their own life and can choose their own course. And this is why, most of the time, feelings of guilt are nevertheless related to moral wrongdoing. Furthermore, to the extent that agency plays a more central role in one's sense of self than one's bodily extension, this also helps to explain why, other things being equal, guilt feelings are bound to be much stronger in cases where agency is involved.

In addition to accommodating cases of traditional wrongdoing, the value-based analysis of guilt also opens the door for some further applications, especially when we consider the constitutive type of contribution. At first glance, it may seem that survivor guilt must stand alone as building on the constitutive relation to a morally bad state of affairs—perhaps because the central value involved in survivor guilt is the somewhat bizarre comparative justice, which seems to differ from other values. Take pleasure or desert as an example. Even though one can cause another person undeserved pain, it does not seem that one can be a constitutive part of another person's undeserved pain because the pain just relates to the condition of the victim. One can cause undeserved pain, but they cannot be a constitutive part of that pain. Comparative justice is different because it attaches to an inherently interpersonal state of affairs, not just to the well-being of the victim. Yet even though the values that give ground to guilt through the constitutive route may be more rare than those that operate on the dimension of causality, they are not limited to comparative justice only. For example, one can also feel guilt for being an integral part of a state of affairs that violates the bodily privacy of another person. Imagine a case of humiliation where one is stripped naked in front of third parties. There is, of course, a sense in which the spectators may feel traditional guilt for not stopping such humiliation from happening, but there is also a special guilt of spectatorship that derives from the fact that it adds to the evil of the situation that the spectator sees the victim naked. Even if the spectator is not the one who does the stripping, their witnessing the victim being stripped adds a further layer of violation of privacy and humiliation compared to a situation where the victim would be stripped in the absence of any third parties.

## 5.2. Comparison with other accounts of guilt

### 5.2.1. A midway between the traditional and interpersonal accounts of guilt

In this section, I want to return to the division between the traditional and interpersonal accounts of guilt, which was one of the driving forces of this dissertation, and explain how the value-based account relates to this division and makes it especially attractive in relation to those two main competitors. As I will argue, the account offers a kind of midway between both types of approaches, combining the best of both worlds. It preserves the outward-looking aspect of the interpersonal account of guilt, according to which guilt is not first and foremost concerned with the self but with what happens out there in the world. Yet, it also preserves the reference to morality which is so deeply embedded in the traditional account. Let me elaborate on both aspects in greater detail.

First, the value-based account of guilt preserves the outward-looking aspect of the interpersonal account of guilt: the main object of concern is not the self but the morally bad state of affairs that one contributes to. Even though I have spent a great deal of this dissertation criticising the interpersonal account of guilt, I do nevertheless agree with one of its basic tenets: that guilt is not so much concerned with the self but with what happens out there in the world. When insisting that “guilt is something that happens between people rather than just inside them”, Baumeister and colleagues underscore the idea that guilt is so much an intrapsychic matter of self-evaluation but is first and foremost concerned with the interpersonal harm that happens out there in the world (Baumeister et al., 1994, p. 243). This basic idea is also shared by several philosophers. In fact, the rise of the interpersonal account was already foreshadowed in Bernard Williams a year before Baumeister’s seminal article. Williams writes:

We can feel both guilt and shame towards the same action. In a moment of cowardice, we let someone down; we feel guilty because we have let them down, ashamed because we have contemptibly fallen short of what we might have hoped of ourselves. As always, the action stands between the inner world of disposition, feeling, and decision and an outer world of harm and wrong. What I have done points in one direction towards what has happened to others, in another direction to what I am. Guilt looks primarily in the first direction, and it need not be guilt about the voluntary. [---] [A]gent-regret [---] can be psychologically and structurally a manifestation of guilt. [---] Shame looks at what I am. (B. Williams, 1993, pp. 92–93)

Williams emphasises that while shame looks at what one is, guilt looks at what has happened to others, and how we have let them down. Since the main focus is on what happens to others, it is not that important whether one was morally culpable or not. This is why he lists agent-regret, an emotion that a lorry driver might feel for running over a child through no fault of one’s own, as a form of guilt. Or at least, this type of outward-looking guilt is the good type of guilt that

Williams talks about—the one that is not just a reverence for some abstract law and the one that does not run the risk of becoming self-indulgent in the sense that the subject would be more concerned with themselves rather than the victims (B. Williams, 1993, pp. 223–224). However, as such, this is also the account of guilt that Williams does not fully develop because, in great part, he is more focused on the critique of the latter type of self-absorbed guilt and the critique of modern moral philosophy in general.

Williams’s idea that guilt looks outwards rather than inwards is also later echoed in Heidi Maibom’s (2019) paper about the difference between shame and guilt. After arguing that guilt and shame are more similar to each other than commonly thought, she concludes her paper with the “hunch” that their main difference lies in their focus: “In shame, the focus is on oneself as someone who has transgressed or failed to live up to certain norms, whereas in guilt the focus is on other persons and how they are affected by what one has done or failed to live up to” (Maibom, 2019, p. 48).

The value-based account of guilt preserves this basic orientation of guilt as an outward-looking emotion that underlies the interpersonal account of guilt and some of its parallel developments in philosophy (which have, unfortunately, not reached their full development as full-blown theories of guilt). I would not say that this outward focus must necessarily concern interpersonal harm (as the interpersonal account does). By speaking in more general terms of morally bad states of affairs, I allow the possibility that the values that underlie guilt need not be confined to the values that are associated with horizontal morality between fellow human beings. The value-based account of guilt can be applied to violations of vertical morality as well, in which case one would view themselves as damaging some higher order, either divine or natural. For example, one might feel value-based guilt about having taken down a grand and beautiful tree because they view it as a violation of the natural order. So, in that sense, the value-based account of guilt is wider than the interpersonal account, and I would suggest that this also makes a stronger and more versatile. But the basic orientation is the same: it looks outward rather than inward.

Of course, by emphasising the outward focus, I do not mean to imply that guilt would altogether lack a reflexive element. It is not an emotion that a mere spectator could feel—the subject nevertheless perceives oneself as contributing to the morally bad state of affairs, simply this contribution does not have to take a morally culpable form. And, most likely, the element of the self is much less pronounced and vague because the attention is not first and foremost on the self but out there in the world.

Secondly, the value-based account also preserves an important aspect of the traditional account—the idea that guilt is a moral emotion. One of the problems with the interpersonal account was that by shifting the focus away from the inward moral evaluation of the agent to the outward harm that befalls the victim, the element of morality went amiss. According to the traditional account of guilt, guilt was a moral emotion because it tracked moral wrongdoing, but the interpersonal account puts the emphasis on a non-moralised notion of harm and thus

does away with morality altogether. Even though the interpersonal account of guilt allows guilt to occur in instances of moral wrongdoing, it is not the wrongdoing aspect that elicits guilt but the harm itself. As I have argued throughout this dissertation, by getting rid of the element of morality, the interpersonal account throws the baby out with the bath water. The value-based account of guilt preserves the deep sense in which guilt is a moral emotion that is clearly and rightly there in the traditional account of guilt, but it expands it by putting the emphasis on moral badness rather than moral wrongness. This is important because even in cases without moral wrongdoing guilt nevertheless reflects a deep concern for morality. In cases of merely causal harm, the subject who feels guilt is concerned about the fact that the harm is unfair toward the victim, and in cases of survivor guilt the subject is clearly concerned with the fact that it is unfair that they made it out alive while others did not. To deny that those cases of guilt reflect a deep concern for morality would be to miss an essential part of what guilt is about.

Showing that guilt can be rightfully called a moral emotion in the traditional sense of the term—as an emotion the core evaluative concern of which involves a reference to morality—is not just a feat within moral philosophy. Acknowledging and addressing the moral dimension of guilt is of crucial importance for those who experience guilt without moral wrongdoing. In Maryann J. Gray’s overview article about the neglected trauma of unintentional killing (which, in great part, concerns cases where one has caused another person to die in a car accident), she notes while those who were only causally responsible do not experience moral injury<sup>112</sup> which goes along with traditional guilt for moral wrongdoing, they nevertheless experience a sort of moral distress (M. Gray, 2021, p. 4). The importance of talking about moral distress in a wide sense of the term has been defended by Campbell and colleagues, who use it to describe negative self-relevant emotions like guilt, shame, lowered self-esteem, anger toward oneself that arise in response to one’s involvement in a situation that is perceived as morally undesirable (Campbell et al., 2018, p. 6). Even though moral distress in this wider sense is not always differentiated from the idea of moral injury, the general idea that many authors seem to share is that people who have suffered a trauma where they have, either culpably or nonculpably, caused someone else to die, it is not sufficient to just treat the posttraumatic stress disorder, which is based on fear, but we need to also address the moral dimension of the experience which manifests itself in guilt and shame (Bica, 1999; M. J. Gray & Yaw, 2024, p. 50; Peters, 2023, p. 36). This all goes to show that there has been some pushback

---

<sup>112</sup> The concept of moral injury is still relatively new and has emerged in the context of soldiers and veterans who suffer from the emotional aftermath of the war. Interestingly, even though the term is best known from the works of Jonathan Shay (2010, 2014) and Rita Nakashima Brock and Gabriella Lettini (2012), it was originally actually coined by a marine veteran and philosopher Camillo “Mac” Bica, who used it in his journals from Vietnam, and has also written about it in academic articles. Moral injury is meant to capture the idea that since the soldiers have had to do things, which, by ordinary standards and their own moral convictions, would be considered as grave moral wrongs, they have sustained a moral injury as in injury to their own moral integrity (Bica, 1999).

against the nonmoralisation of cases without moral wrongdoing. As also noted in Peters's (2023) autobiographical self-help book: there is a need for language that would help us make sense of the "murky moral area" (p. 36). For cases where there is no moral wrongdoing, I hope the value-based account of guilt can explain how exactly it is that a profound sense of morality nevertheless enters into guilt, even in the absence of moral wrongdoing.

To sum up, the value-based account of guilt is an attractive account of guilt because it preserves, though perhaps in a somewhat modified form, some of the best features of both the traditional and the interpersonal account of guilt. First, from the interpersonal account of guilt it takes the idea that guilt is an outward-looking emotion—the idea that has also been shared by philosophers like Williams and Maibom—but rather than defining this outward focus in the narrow terms of interpersonal harm, it puts the emphasis on morally bad states of affairs more generally. Secondly, there is also a kernel of truth in the traditional account of guilt that the value-based account had sought to preserve: the idea that guilt is not just nonmoral sensitivity to harm but a deeply moral emotion. But rather than defining the moral aspect in the narrow terms of moral wrongdoing, it puts the focus on the more general idea of moral badness. As I have argued, the interpersonal account of guilt throws the baby out with the bathwater by altogether losing the reference to morality in their definition of guilt as a matter of nonmoral harm. That this has proved problematic is also shown by the fact that there is a pull towards the remoralisation of guilt among the people who work closely with those who feel guilt for unintentional killings. Practice shows that their trauma cannot be fully addressed if its moral aspects are ignored. However, within the psychological literature, the resources for accommodating this moral aspect, especially in complicated cases without moral wrongdoing, are scarce. But as I have suggested, going back to the fundamentals in normative ethics, that is, to the division between the good and the right, and saying that guilt is not necessarily about moral wrongdoing but about moral badness, offers us a simple way to disentangle those complicated situations.

### 5.2.2. Comparison with other value-based accounts of guilt

Now that I have restated what the value-based account preserves and leaves out from two of its broader rivals, the traditional and the interpersonal account of guilt, I would like to further clarify the value-based account of guilt by contrasting it with some of its closest relatives on the philosophical landscape. In particular, I will focus on two alternative accounts—Gabriele Taylor's (1985) account of guilt as a matter of responsibility for the forbidden, and Michael Zhao's (2020) self-implication account of guilt. Let it be said that there are also differences in how we frame our theories—while Taylor and Zhao defend cognitive theories of guilt, I remain neutral with respect to emotion theory. But, as I have already argued in section 2.4, different theories of emotion do not prevent us from having meaningful discussions about the core relational theme of guilt. So, the differences that I will touch upon concern our construal of the core relational theme.

Let us begin with Gabriele Taylor's (1985) idea that guilt is a matter of being responsible for something that is perceived as forbidden. Her talk about the forbidden state of affairs is sufficiently similar to my talk about moral badness, but let us consider the other aspects of her account. Taylor operates with "the widest reading of 'responsibility'" where it can be merely causal responsibility as well. She does emphasise that a driver who kills a child through no fault of their own is causally responsible for a terrible (forbidden) thing, and that this is sufficient for guilt (1985, p. 91). As such, her idea of responsibility does capture the same extension as my notion of causal contribution. But, as we already saw, she had difficulties recognising that one can also contribute to morally bad states of affairs in ways other than the causal (remember the somewhat bizarre way how she tried to make guilt for one's privileged position a matter of causing this by one's birth (Taylor, 1985, p. 91)). And secondly, for Taylor, the bad state of affairs is not the primary focus. She stresses: "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.'" (1985, p. 92) For this reason, she ties guilt with perceiving oneself as stained (1985, pp. 92, 94, 100, 103). The view that I am suggesting is the opposite: the subject is more concerned with what has happened out there in the world and the phenomenology of the emotion is better captured not in terms of having a stain on oneself but in terms of having stained the world—having a negative moral footprint.

Secondly, even though our accounts have been developed independently, Michael Zhao seems to reach similar conclusions in his (2020) article that focuses on guilt for merely causal harm. Just to be clear, since the author has defended "a novel account of guilt" twice (Zhao, 2020, p. 286; MacKenzie & Zhao, 2023, p. 2708), it should be highlighted that this is a different account of guilt than the one by MacKenzie and Zhao that was discussed in the survivor guilt chapter (the account that took a Vellemanian turn, which I criticised as a misguided step backwards). In his 2020 article, Zhao offers a brilliant analysis of guilt for harm to others, which puts a great emphasis on the fact that the subject has caused a morally bad state of affairs. In that sense, his account is a value-based account as well because he operates with evaluative rather than deontic terms (even though he does not explicitly elaborate on that). The main difference between our accounts comes from how we explain the relationship between the self and the morally bad state of affairs. Whereas I speak of contribution, Zhao speaks in terms of *implication*, in fact, he even calls his account a self-implication account of guilt (2020, p. 307). According to this account, "[g]uilt represents some part of the self as being implicated in a bad state-of-affairs" (p. 307). And the way he defines the relationship of implication is the following: "When a capacity produces certain objects of a kind production of which is its function, in the way that our volition produces intentions, I'll say that it is *implicated in* the existence of those objects" (p. 309). So, what is important, is that some capacity of ours that we identify with would produce a morally bad state of affairs.

Since my aim in this section is just to contrast the main content of my own account and that of Zhao's, I will only focus on the comparison between Zhao's

account of implication with my own account of contribution. It is also useful to note that our accounts differ with respect to the fact that Zhao proposes his account as a theory of all guilt, whereas I just present it as a sense of guilt that arises from a value-based moral outlook. As I will later go on to argue, any value-based account of guilt that puts the focus on moral badness rather than moral wrongdoing is bound to have difficulties accommodating some forms of guilt that arise from a deontological moral outlook. But this will be the topic of 5.3.2.

In order to get a better grip on Zhao's central concept of implication as a matter of some capacity producing certain objects, the production of which is its function, let us see how his account applies in all the crucial types of cases: traditional cases of wrongdoing, cases of merely causal harm and survivor guilt. Let us first look at cases of moral wrongdoing. As a background condition, it is important that the subject identifies with their volition, which is a capacity to produce intentions. In the case of moral wrongdoing, guilt is felt about the fact that one's volition has produced a morally bad intention. (In section 5.3.2, I will argue that some deontologists would insist that the intention is morally *wrong*, not first and foremost or even necessarily *bad*. But for the current purposes this does not matter. Zhao's account will nevertheless apply to some instances of moral wrongdoing—those that arise from a value-based moral outlook, so let us keep the focus on that.) Next, there are cases of merely causal harm, which Zhao treats in two parts. First are cases that, even though they do not involve moral responsibility, nevertheless involve agential responsibility. As a background condition, the subject identifies with their agency, the capacity to produce actions. In the case of merely agential responsibility, one's agency produces a morally bad state of affairs, which, for example, in cases of car accidents, is constituted by the subject killing a pedestrian. Secondly, we have non-agential cases of merely causal harm, for example, those where one is just instrumental to harm (imagine the case of Elvis absorbing more than his fair share of nutrients in the womb, so that his twin dies, or the cases where one is used as an instrument of evil and thrown on top of others). As a background condition, the subject identifies with "their causal presence, that is, their capacity to stand in causal relations to other things, to affect and be affected by other parts of the world" (Zhao, 2020, p. 311). In this case, their causality produces a morally bad state of affairs (the death of another person). While his 2020 article is not really focused on survivor guilt, Zhao does emphasise that the account is open-ended enough to be extended to survivor guilt as well. He offers the following, admittedly "very sketchy" account of survivor guilt towards the end of his article:

[I]n the case of survivor guilt, perhaps we identify with our capacity to stand in relations to others, regarding that capacity as part of the self. To the extent that these relations are ones of undeserved inequality (as when we survive something that others do not), we might think that they are to that extent bad, and our relational capacity becomes implicated in the existence of something bad. (Zhao, 2020, p. 314)

There is much that can be said about Zhao's ideas on implication, but there are two problems in particular that I would like to highlight: it both underestimates and overestimates the role of causality. First, at least at an explicit level, it underestimates the role of causality because Zhao only refers to causality in cases where one is merely instrumental to harm—in those cases, one identifies with their causal presence, and their causal presence produces bad outcomes. In other cases, the subject identifies with something else. But, as I hope that my inquiry has revealed, causality actually plays a much bigger role. In fact, Zhao's very notion of 'production', which he refers to in his definition of implication (“[w]hen a capacity produces certain objects of a kind production of which is its function, [---] I'll say that it is *implicated* in the existence of those objects”) is in great part a causal matter. For example, if volition is said to produce intentions, how else is that supposed to happen than not without causally bringing those intentions about? Also, as already discussed, our accounts of agency (including moral agency) are usually built on the causal theory of action, according to which actions are events that are caused by our mental states like intentions. So, we cause our intentions, and our intentions cause our actions. All those forms of 'production' are implicitly causal matters. Yet causality, at least if we adopt the counterfactual understanding, covers more than production, which could at best be associated with the more naïve view of causation, according to which there has to be some physical (or otherwise special)<sup>113</sup> connection between the cause and the effect. Talking in terms of production does not seem to enable us to make sense of why people can also feel guilt for omissions: we do not normally say that we produce the results or our omissions. But for the purposes of talking about guilt it is essential that a good account of guilt could accommodate why people can feel guilt for omissions as well. Overall, for both clarity and better coverage, it would be better to substitute the vague talk about production with a counterfactual understanding of causality.

Secondly, to the extent that causality seems to be implicitly built into Zhao's notion of production, it can also be argued that it does not seem to leave enough space for constitutive contribution. Zhao says that, in case of survivor guilt, our capacity to stand in relations to others is implicated in the relation of undeserved inequality, but it is not clear in what way we can say that our relational capacity 'produces' this unjust state of affairs. As can be seen, on closer inspection, much seems to hang on Zhao's notion of production, and it is indeed unfortunate that he gives no further clarification with regard to what it means to produce something. If he wants to make his notion of production work in the right way, it seems that he would ultimately have to fall back to the account that I have been suggesting: the subject must *contribute* to a morally bad state of affairs in the sense that the morally bad state of affairs has to be counterfactually dependent on the subject, and the two ways of contribution are the same old causal and constitutive

---

<sup>113</sup> It is not clear whether the way how volition causes intentions can be described in physical terms.

relations that (differently from ‘production’) have figured in philosophical discussions for a long time. As such, I take it that I have managed to capture the relationship between the subject and the morally bad state of affairs in more clear and fundamental terms than Zhao’s somewhat vague account of implication in terms of production.

### **5.3. Normative ethics and the limitations of the value-based account**

Especially for those who are sympathetic towards the possibility of guilt without wrongdoing, it might be tempting to think that all guilt should be understood in terms of the subject contributing to a morally bad state of affairs. This is related to the fact that those people are likely to have a value-based moral outlook, which is the type of moral outlook that gives rise to those kinds of guilt feelings. But it is important to remain vigilant about the danger of imposing our own moral outlook on a general conception of guilt that is supposed to capture how all people experience guilt. In fact, this seems like one of the greatest pitfalls in formulating an account of guilt—blindness to the fact that all people might not experience guilt in the same way as we do. This mistake is evident in many accounts of guilt. The interpersonal account of guilt that speaks of guilt in terms of harm to others, and thus focuses on horizontal relations, turns a blind eye to the fact that in many cultures that are not categorised as WEIRD (Wester, Educated, Industrialised, Rich and Democratic) people report guilt for religious violations of more vertical nature which consist in disobedience to some higher order, and as such their understanding of morality is not easily reducible to the kind of horizontal morality that governs the relationships among equals.<sup>114</sup> The mistake is also evident in what I have called the traditional account of guilt that supposes that people perceive guilt first and foremost in terms of moral wrongdoing, and thus excludes the possibility of guilt without wrongdoing. This dissertation has been focused on correcting this mistake—showing that, for many, it is moral badness rather than moral wrongness that is of central importance. But thinking that the new, supposedly wider account of guilt, should now be proposed as an account of all guilt would be to commit a similar folly—to think that all people must necessarily perceive questions of morality in terms of moral badness. This mistake has been committed by Michael Zhao’s 2020 version of a value-based account.<sup>115</sup> My aim, however, is to avoid this mistake. So, in this section, I want to be especially clear

---

<sup>114</sup> This criticism is inspired by Francesco Mancini and Angela Gangemi’s account of (what they call) ‘deontological guilt’, which is in great part a matter of playing God / tampering with Nature (Mancini & Gangemi, 2021).

<sup>115</sup> Even though he calls his account differently (as the self-implication account of guilt), his account, too, is a version of a value-based approach to guilt because he, too, shifts the attention from moral wrongness to moral badness, even though he does not elaborate on that shift in much detail.

about which types of moral outlooks and theories the value-based account of guilt does resonate, and with which it does not. After explaining the relationship between moral outlook and moral theories, I will first clarify the types of moral theories that are compatible with the value-based account of guilt, and then move on to discuss which are not, and why the account cannot serve as an account of all guilt.

### 5.3.1. On the moral theories that the value-based account does cover

The value-based account, a version of which I have defended in this dissertation, arises from a value-based moral outlook, by which I mean a way of perceiving moral matters that puts the main focus on the moral badness of states of affairs rather than the wrongness of one's actions. Most importantly, it puts the good before the right and assesses the rightness and wrongness of actions based on whether they bring about good or bad states of affairs (broadly construed, so that the performing of an act itself may constitute a state of affairs as well). As described in 5.1.1, I do not think that the value-based moral outlook should necessarily take the form of having a full-blown moral theory. Also, I do not think that endorsing a certain moral theory will necessarily mean that one will have the relevant moral outlook, i.e. that their lived moral experience, including patterns of attention and emotion, necessarily correspond to one's moral theory. But I am still optimistic that, ideally, the two would go hand in hand, and would thus like to clarify with what kinds of moral theories the value-based moral outlook resonates with.

So far, I have knowingly avoided speaking in terms of consequentialism, deontology and virtue ethics. Following (Vallentyne, 1987) and (Hammerton, 2025), I do not think that those textbook categories capture a neat fundamental division between normative theories. According to Hammerton, they are "historically loaded family resemblance concepts" (Hammerton, 2025, p. 337), which are themselves crisscrossed by other, more fundamental divisions like those of agent-relativity/agent-neutrality, normative priority, and maximisation, which do not really enable us to divide theories into two discrete categories of consequentialism and deontology but result in a far more rich landscape of options that may differ along many dimensions. This is precisely why I have organised my talk about guilt around the more precise contrast between value-based versus alternative moral outlooks, which pertain to the dimension of normative priority.

I do want to clarify, though, at least in an exemplary manner, which of the more specific theories those are that reflect a value-based moral outlook. While there is little debate about consequentialism being value-based, it is useful to note that some theories that are commonly described as belonging to the sphere of deontology and virtue ethics can have a value-based structure as well. Admittedly, for some, this may well be a reason why we should not call those theories *truly* deontological or *truly* virtue ethical, but this is a debate I want to side-step because, as already said, I am sceptical about the prospects of a single rationale that would

enable us to neatly individuate consequentialism, deontology, and virtue ethics without any overlaps and grey areas. So, I will be using those labels only in a loose and not in a particularly pretentious sense.

Let us start with the obvious: it is relatively safe to say that consequentialist theories are value-based because they put the good before the right—they start by attributing value to certain kinds of states of affairs and then work their way to determine which actions would bring about such states of affairs.<sup>116</sup> Hedonistic utilitarianism, according to which we should bring about pleasure and reduce pain, is a classic example. The basic line of thought is the following: it is right to promote pleasure because pleasure is good, and it is wrong to inflict pain because pain is bad.

Since the priority of the good is such a classic feature of consequentialism, let me be extra clear about why I prefer *not* to call my account consequentialist, though. For many theorists, consequentialism is not just a matter of focusing on valuable states of affairs but also involves other features, most notably, the idea that one must relate to those valuable states of affairs in an agent-neutral manner (McNaughton & Rawling, 1991; Howard-Snyder, 1994; Pettit, 1997). In other words, this means that the states of affairs that you should promote or refrain from do not depend on your position as an agent—everyone should aim for the same general good. Whereas agent-relative normative theories emphasise the idea that, as an agent, you may have a special reason to do some things that others do not (e.g. take care of your children rather than someone else's), and to refrain from certain actions like lying or killing, even if the particular lie or killing would serve a good end and would help to reduce the overall amount of lyings and killings in the world. Admittedly, critiques of consequentialism on the grounds of agent-neutrality have brought about theorists who identify as consequentialists but defend agent-relative accounts of value (Sen, 1982; Broome, 1991; Portmore, 2001, 2003, 2011), and there are also other authors who prefer broader definitions of consequentialism that do not require agent-neutrality (Bennett, 1989; Broome, 1991, pp. 5–6; Skorupski, 1995).<sup>117</sup> Yet I doubt that, for a person who is not familiar with those debates, an agent-relative construal of consequentialism might come to mind at all. So, as can be seen, people have different associations with consequentialism, and what counts as true consequentialism is a contested matter. I do not intend to settle those debates. It is simpler for my purposes to just go

---

<sup>116</sup> In his overview of consequentialism, Walter Sinnott-Armstrong notes the possibility that, strictly speaking, consequentialism need not take an evaluative form. Technically, we could construct a consequentialist theory that assesses the rightness of action without any reference to value but simply based on whether it increases the number of goats in Texas. Admittedly, such a theory may be absurd, but its basic shape would still be consequentialist in the sense that we are focused on states of affairs. (Sinnott-Armstrong, 2023, p. 6) For the purposes of the current inquiry, I have chosen to ignore this possibility and focus on consequentialism in the more narrow evaluative form that is not just a theoretical possibility but the form in which consequentialism is usually proposed.

<sup>117</sup> More about those theories and debates can also be read from (Sinnott-Armstrong, 2023, pp. 5–6, 23–25).

straight to the point and talk about the feature that matters for my account—the idea that the normative dominant must be the value that is attributed to states of affairs. Whether one is oriented toward the good within an agent-neutral or agent-relative frame is not of crucial importance. For example, let us imagine that a person feels guilt about the fact that they have compromised their children’s well-being. If what they lament the most about the situation is, nevertheless, the fact that they have brought about a bad state of affairs, then it does not really matter whether one relates to the relevant values (the child’s well-being) in an agent-neutral or agent-relative manner. Within both frames, one’s guilt can nevertheless be oriented towards the bad state of affairs. (Of course, agent-relativity need not go hand in hand with placing the good before the right, but I am here only talking about the cases where it does.)

Moving on, some construals of deontology, too, take the same value-based form. For example, people often speak of Kantian ethics as having a value-based structure. I am not saying that this is necessarily the right way to interpret Kant, but if we interpret Kant’s dictum to never use human beings as mere means to our own ends but always as ends in themselves as a matter of respecting autonomy as an important value, then the basic form of such account would be value-based. The starting point would be to acknowledge the value of autonomy—of people being able to decide for themselves about matters concerning their own life—and the next step is to say that those actions which honour this value are right, and those that do not are wrong. Christine Korsgaard (1996), in particular, has been said to “appear to attribute this structure to Kant’s moral theory” (Hammerton, 2025, p. 325). It is not my aim to ascertain whether this is a fair interpretation or not. Neither is it my aim to settle whether a theory with this kind of structure should rightly be called deontological. Emphasising the importance of autonomy does indeed reflect a more victim- rather than agent-centred brand of deontology, which is known to run into the trouble of consequentialising deontology (Alexander and Moore, 2024). Some proponents of deontology suggest that such contemporary talk about autonomy as a valued state is consequentialist (Gaus, 2001, p. 182), and some authors who list autonomy as a valuable good do also explicitly identify their theories as consequentialist (Griffin, 1986). But regardless of how exactly we should label a value-based approach to autonomy, it is nevertheless clear that what many people are used to describing in terms of Kantianism can nevertheless have the kind of value-based structure that is relevant for value-based guilt.

In fact, there are authors who have even gone as far as to attribute value-based structure of deontology as a whole. This applies to Philip Pettit’s construal of deontology (1989), which he offers for the purposes of demarcating between consequentialist and deontological theories. Pettit suggests that even though both can start from the same theory of the good, consequentialist theories seek to promote the good, while deontological theories aim to honour the good. We get a clearer idea of what this means when we look at how he illustrates the idea on the example of being peaceable:

With a value like that of being peaceable, I promote it if I do whatever promises the maximal realization of the value; this may include not being peaceable myself, as in fighting the war to end all wars. I honor that value on the other hand if I choose options that exemplify it, being peaceable myself, even if this means that there is less peace overall. (Pettit, 1989, p. 117)

So, the main difference lies in the idea that while the consequentialist seeks to make the world as peaceable as possible (even if that comes at the expense of not being peaceful themselves), the deontologist is mainly just concerned with being peaceable themselves. The important point for our purposes is that, in both cases, moral theorisation starts from a theory of the good, and in deontology, too, the right is derived from the good—the right is that which honours the good. Admittedly, Pettit (himself a consequentialist) has henceforth been heavily criticised for his depiction of deontology (McNaughton & Rawling, 1992),<sup>118</sup> but leaving that as it may, the point I want to emphasise by drawing attention to Pettit’s construal of deontology is that, what many people are used to calling deontology, fits quite well with the value-based moral outlook. It definitely resonates with the textbook idea, according to which deontology is a matter of actions being intrinsically right or wrong, regardless of the further consequences (besides the taking place of the action itself) that they may bring along.<sup>119</sup> And to the extent that deontology is often depicted as a matter of honouring a value, especially autonomy, in the sense of not bringing about violations of that value, then this will be relevant for the value-based sense of guilt because the main focus of the person who feels guilt will be on the negative state of affairs that consists in violating the victim’s autonomy.

Furthermore, my account of guilt can also resonate with at least some forms of virtue ethics.<sup>120</sup> For example, Christine Swanton (1995, 1997) has defended a

---

<sup>118</sup> As pointed out by McNaughton and Rawling, Pettit’s construal of deontology is both historically inaccurate in that it “seriously distorts a main strand in deontological thinking” and leaves deontology overly vulnerable to the kind of consequentialist attack that they refer to as the consequentialist vacuum cleaner (1992, pp. 836, 842). What they are talking about is the consequentialist suggestion that if value is placed on states of affairs, then there is no rationale for arguing in defence of agent-based restrictions—if my doing a small unjust thing would help to promote a greater level of justice out there in the world, then there is no rationale for me to not do the unjust thing. As such, value-based forms of deontology are vulnerable to getting sucked into the consequentialist vacuum cleaner. For readers who are perhaps familiar with John Harris’s classic text “The Survival Lottery” (1975), this is a perfect example of the consequentialist vacuum cleaner in action: as Harris points out, if what we value is human life, then, if by killing one, we could save the many, then the consistent thing would be to save the many.

<sup>119</sup> On the problems of this textbook view of deontology, see (Vallentyne, 1987, p. 23).

<sup>120</sup> For the sake of simplicity, I am ignoring here the distinction between ‘virtue ethics’ and ‘virtue theory’ (Driver, 1996). According to this division, only those theories can be counted as truly virtue ethical that give virtue a central place in the normative structure of the theory, whereas those accounts of virtue that are embedded in a different normative structure are called ‘virtue theories’. See also (Hammerton, 2025, p. 332).

value-centred approach to virtue ethics, which construes virtue as a habit of responding appropriately to value. She proposes her approach as an alternative to the neo-Aristotelian idea, according to which virtue is a character trait that helps the agent flourish. Swanton stresses that such construal takes virtue too far from right action, and does not provide us with a rationale that should motivate us in our daily dealings—if an action is meant to promote one’s personal flourishing, then “such promotional behaviour may actually disqualify the action as one exhibiting the virtue in question” (Swanton, 1997, p. 168). So, instead of saying that virtue is a disposition that promotes agent flourishing, she finds it much more attractive to speak in terms of values that provide a direct rationale for each virtue. For example, benevolence is a matter of valuing other people’s welfare, whereas justice is a matter of valuing justice. Swanton herself is keen to emphasise that virtue is not just a matter of promoting value, but also a matter of honouring value, and furthermore, can even involve further modes of response like expressing and appreciating. Also, even though one can respond to the same value in different ways, different virtues nevertheless have more characteristic modes of responding to the relevant value. For example, benevolence is characteristically a matter of promoting, for example, the welfare of other people, whereas the virtue of justice is characteristically a matter of honouring justice by being just oneself and not dirtying one’s hands in unjust deeds (Swanton, 1995, p. 50). I do not think that expressing and appreciating are related to *contributing* to valuable states of affairs, which has been the focus of the type of guilt I have been discussing. This is because expressing and appreciating are first and foremost matters of having the right attitude. They are not first and foremost related to making a difference to the world. But to the extent that Swanton’s account involves honouring and promoting, those modes of response are undoubtedly related to contribution, because when one promotes a value, they bring about valuable states of affairs, and when one honours a value, they avoid bringing about disvaluable states of affairs. Again, I suspect that some might be reluctant to call Swanton’s account true virtue ethics and might prefer to say that she simply offers a theory of virtue from a broadly consequentialist framework.<sup>121</sup> But it nevertheless goes to show how at least some people who identify as virtue ethicists may nevertheless rely on a value-based framework. And for them, the kind of guilt that would be elicited by either failing to promote or respect a value is likely to take the form that I have described—it will involve a painful awareness of contributing to a morally bad state of affairs.

Overall, I hope to have shown that, when we look at the landscape of normative theories, we can see that it is not only consequentialism that has value-based structure, but that value-based structure can also be a part of what at least some theorists have described as deontology and virtue ethics. All those who adopted

---

<sup>121</sup> Perhaps not unlike Thomas Hurka’s (2000) consequentialist account of virtue as a matter of loving the good and hating the evil. There are also direct links between Swanton’s account and Hurka’s. Swanton refers back to Hurka’s (1992) idea of the virtuous person loving the good in her (1995, p. 59).

the above-mentioned theories as live moral outlooks—that is, all whose emotional life and patterns of attention are attuned to valuable states of affairs as the normative dominant—are susceptible to feeling value-based guilt that focuses on the self as standing in a contributive relation to a bad state of affairs.

### 5.3.2. On the moral theories that the value-based account does not cover

Next, let us move on to the kind of approaches that do not have a value-based structure. Strictly speaking, theories that derive the right from the good do not have a single opposite. It would be a mistake to contrast them as theories that just reverse the order of priority and derive the good from the right because this does not exhaust the possibilities—one can also make the right and the good independent of each other (Vallentyne, 1987, p. 24). All this leaves room for a great multiplicity of non-value-based theories, and I do not have the pretension of giving a comprehensive account of all of them. The main focus of this dissertation has been on showing that the kind of guilt that is perceived in cases without moral wrongdoing is not just some loose way of talking about guilt but can be explained by a rather coherent value-based framework. Given this focus, I could hardly be expected to give a complete account of all the other ways how one might experience guilt. However, I will aim to illuminate the main line of thought behind at least some of the most prominent alternatives. This will be sufficient to serve as a counterexample to show that the normative landscape is not exhausted by value-based accounts of morality, and thus the value-based account of guilt, too, cannot be an exhaustive account of guilt. In addition, this will also help us to understand why some proponents of the traditional view of guilt have chosen to speak in terms of the agent perceiving *oneself* as morally stained (Ingram, 2017), which stands in stark contrast with how those who experience a value-based form of guilt might describe their experience in terms of leaving a negative moral footprint on the *world* or violating the moral order of the *world*.

An especially useful starting point can be found in Stephen Darwall’s (1986) distinction between those normative theories that work from “the outside-in” and those that work from “the inside-out”. Even though Darwall himself is concerned with moral theory rather than guilt feelings, framing the matters in terms of outside-in and inside-out is especially useful for understanding the phenomenology of moral experience, including guilt. Value-based accounts that we have been talking about work from the outside in: they start from attributing intrinsic value to certain types of states of affairs out there in the world, and then work their way in—to right action, and to right character, which both will be derived from the kinds of states of affairs that they are conducive to (Darwall, 1986, p. 308). The proponents of the alternative view are often deeply sceptical about the prospects of defending an objective account of the good (O’Neill, 1996, Chapter 2; Korsgaard, 1996), and do not find it as a particularly good starting point to ground our moral duties. Instead, it begins inside the moral agent, with

the idea that everyone is responsible for their own moral integrity (Darwall, 1986, p. 311), or so-to-say for keeping their own moral house in order (Alexander & Moore, 2024, p. 8). This is what Darwall suggests ultimately explains the importance of agent-based restrictions (the idea that I have a special reason to refrain from certain forms of actions, e.g. killing, even if my performing those actions would serve the purpose of reducing the overall number of such actions), which he and many others find particularly definitive of deontology (Darwall, 1986, p. 295; Gaus, 2001, p. 185; McNaughton & Rawling, 1991, p. 168, 1998, p. 39). The focus will be, not on which kinds of states of affairs one should produce, but on the principles that one should be guided in their deliberations about choice of acts (Darwall, 1986, p. 312). For example, with the first formulation of the Categorical Imperative, Kant emphasises the importance of consistency—we are only supposed to act on maxims for which we can, without any contradictions, will that they become a universal law. For instance, lying does not pass that test because if everyone else lied, lying would become an ineffective means to achieve the end that the subject is aiming at—no one would believe the liar. As such, the liar just wants to free-ride a system, the existence of which is necessary for them to succeed with their plan. The principle of universalisability is known to be formal in the sense that it does not appeal to any substantive account of the good.

To some extent, this poses a problem for the respective account of guilt feelings. Emotion theory is known to rely on a conceptual tie between emotions and values (including anti-values). When we experience an emotion, this emotion shows something about what it is that we value, that we care about. In that sense, even the traditional definition of guilt as a matter of moral norm violations makes guilt stand out as somewhat odd among other emotions.<sup>122</sup> It seems that a norm alone, without any respective value that it would be linked with, is not sufficient to elicit emotion. The value-based approach solves this problem because norms are seen as derived from the inherent value that is placed on certain states of affairs. So, when one feels guilt, they feel guilt for the fact that they have, for example, damaged something of value. If, however, morality is supposed to be a matter of pure rationality, then it is unclear how it can elicit any emotion at all. But one way to solve this seeming mystery is to go back to the starting point that Darwall mentions—the idea that everyone is first and foremost responsible for their own integrity. Darwall speaks of this as grounding the second-order duty not to compromise one's integrity. This duty, at least, could be viewed as arising from the value that we place on our own integrity. So, what a Kantian would have to be concerned about, when they felt the emotion of guilt, would have to be the fact that they have compromised their integrity.

---

<sup>122</sup> That guilt lacks a direct connection with values becomes especially salient when Fabrice Teroni and Julien Deonna note in their comparison of shame and guilt that the former is related to values (whether one exemplifies the self-relevant values or ego-ideals), and the latter with norms (Teroni and Deonna 2008, 732).

This is indeed a starkly different way of talking about both morality and guilt feelings. For a person with a strongly value-based moral outlook, this may sound incomprehensible, but when we look at how some people talk about guilt, then this is indeed precisely the case. Let us consider Andrew Ingram's "moral staining view of guilt" (2017). Ingram is a proponent of the traditional, deontic view of guilt, and emphasises that since people also report guilt and scrupulousness in case of misdeeds that involve no substantial harm to the world, it would be better to define guilt by focusing on its reflexive, inward-looking character. One of his central examples concerns a scrupulous student Emily. Emily is doing her geography exam and trying to remember the name of one country, but, before the right name comes to her mind, she accidentally catches a glimpse of the right answer because a book which she had earlier carelessly crammed into her backpack was accidentally open from the right page, and the answer was peeking out through her backpack. Thinking that the name of the country was at the tip of her tongue anyway, she nevertheless decides to write it down. Later, however, she becomes scrupulous about whether what she did was wrong:

'No, what I did was wrong. But, does that mean that I am a cheater? I've never done anything like that before, and I won't do it again... I hope... but there's still some bad in me that made me do that. Can I really truthfully say that I'm an honest person, honest deep down?' (Ingram, 2017, p. 11)

It is clear that Emily's scruples have little to do with having made the world into a worse place, she is concerned with her own moral character. According to Ingram, this exemplifies the concern that one may have about moral wrongdoing leaving an "ethical black mark" on one's identity. Relying on a Korsgaardian (2009) view according to which we constitute our practical identity through our chosen actions, moral wrongdoing leaves a mark on our identity. Our track record is not clean anymore:

When we choose to act unethically, we are like craftsmen inserting rotten timbers into a frame house. Our work is thus sullied in the same way that the work of a cook who pours a spot of rancid milk into the soup mars his own labors. Like the soup, the self is a work-in-progress to which many ingredients have been added. One does not become a wicked person by one wicked action, but like the soup, one nonetheless has at least one undeniably bad element composing oneself. The ethical stain that bad actions leave is just their contribution, a few shoddy nails and boards, to the ongoing process of self-construction. When we choose to do wrong, we reshape ourselves, however so little, for the worse. (Ingram, 2017, p. 7)

The reason why I cite this paragraph in full is that metaphors do often give us a better phenomenal grasp of what it is to experience a certain kind of emotion, and here, the metaphors around which the moral staining view of guilt is built around are starkly different from the ones that we have seen in the context of value-based guilt. A person who experiences value-based guilt might describe their experience in terms of their moral footprint in the world, or they may view their survival as violating the moral order, as making the world into a more just place.

Here, however, the concern is not with the mark that one has left on the world, but on the mark that one's action has left on oneself. This exemplifies a completely different moral orientation—the focus is not on what happens out there in the world but on our own integrity, on whether we have managed to keep our own house in order.

By pointing at Ingram's moral staining view of guilt, I do not intend to suggest that someone with a value-based outlook cannot be concerned about their integrity and practical identity. Of course, they can. After all, Ingram proposes the moral staining view as a general account of guilt, and as such, it would make little sense if people with a value-based moral outlook were utterly incapable of being concerned about their moral integrity. He is simply concerned that a more harm-based view does not capture the cases of harmless wrongdoing like that of Emily's, and suggests that we must thus put the focus on the reflexive character of guilt rather than the outward harm. But the point is that such an account is especially catered to capture the moral experience of someone with a non-value-based moral outlook, of someone, who, like Darwall suggests, moves from the inside-out and starts from the idea that one is first and foremost responsible for one's own integrity. And the more general idea that my dissertation points at is that just like the value-based account of guilt cannot make sense of guilt as experienced from those more agent-based forms of deontology that have an inside-out rather than outside-in logic, the moral staining view, which is specially catered to do justice to those experiences comes at the expense of not giving a particularly convincing account of the guilt that is experienced by those with a value-based moral outlook. Because, for them, morality does not start from the fundamental responsibility for one's own integrity. As such, it is highly unlikely that concern for having left an "ethical black mark" on themselves (if present at all) would be the most salient feature of their guilt experience. When a person with a value-based moral outlook feels guilt about wrongdoing, then what wrongdoing means for them is the fact that they make the world into a worse place, and it is this fact that they will be first and foremost concerned about.

This dissertation was born from the desire to make better sense of the guilt feelings of those who report guilt without perceived wrongdoing. However, as the latter discussion has shown, it is not only that people with the deontic view of guilt are likely to misunderstand those who experience value-based guilt. If we pull the deontic view to an agent-based extreme where it cannot bleed out of the bounds of the deontic, then it becomes evident that those with a value-based moral outlook are also very likely to misunderstand the guilt feelings of those who view guilt through the moral staining view. I suspect that most of the time, this misunderstanding remains hidden because many instances of moral wrongdoing are such that they can be explained through a value-based lens as well. However, when we look at Ingram's example of Emily, it should become evident that for a person with a deeply value-based outlook, this example is likely to remain as incomprehensible as the cases without moral wrongdoing to are likely to remain incomprehensible to those with a narrowly deontic view. But the fact that those differences only come to the fore in the extreme ends of the guilt spectrum does

not mean that they are not also there in the regular cases as well. If proponents of different moral theories feel guilt about stealing, for instance, their emotional experience is likely to be very different—the one with a value-based moral outlook will be focused on how they have caused unfair damage to the other person, whereas a proponent of agent-based deontology is likely to be concerned about how *they* could have done such a thing. This also has implications on what it is that will help the person to overcome guilt. The value-based view will push us towards repairing the damage we have done to the world, whereas other views are going to push us towards working with the self.

If we thought that guilt is a matter of moral norm violations, as the traditional account of guilt would suggest, then it would seem that guilt feelings mainly just vary with respect to what it is that people consider as morally wrong. This leaves the basic shape of guilt feelings intact—guilt is always felt about moral wrongdoing, only what people consider as morally wrong may vary. But what people consider as right or wrong is just a symptom of their broader moral outlook. When we delve into why it is that people consider certain actions as right or wrong, we can see that there are many different dimensions along which deep conceptual disagreements arise (Hammerton, 2025; Sutrop, 2020). In this dissertation, I have focused on the dimension of normative priority because it seems to me that compared to other divisions that might concern, for example, agent-relativity versus agent-neutrality, maximalisation, differing values and conceptions of good life, normative priority is the one that has the most dramatic impact on the very shape of guilt feelings. This is because the reason why we consider certain actions as right or wrong is going to have a direct impact on what it is that we are focused on when we experience guilt. If our normative dominant is the good, then we are bound to be focused on how we have contributed to a morally bad state of affairs. If, however, we are more concerned about the features of our own will and character, then our attention is more likely to turn inward rather than outward. As it can be seen, this has a dramatic effect on not just what kind of actions we might feel guilt about but on the very shape of guilt feelings: the value-based account of guilt is oriented towards how one has contributed to what happens out there in the world, whereas more agent-based accounts can lead us to the moral staining view of guilt where the focus is on how the wrong action has stained the self rather than damaged the world around us. In her (2020) article on moral disagreements, Margit Sutrop emphasises how, in the case of deep conceptual disagreements about morality, alternative perspectives are conceptually unavailable to the disputants. My work on guilt feelings shows how this carries over to our theorisation of guilt. For those with value-based moral outlook, it can be very hard to understand how guilt could be conceptualised as a matter of moral staining of the self, whereas for proponents of more agent-based deontology, it may seem largely irrelevant to be mainly focused on what happens out there in the world. Neither of those views is going to strike a chord with all people. As such, it is crucial that, as we try to construct theories of guilt, we remain vigilant about how our own moral outlook can influence our theorisation and not try to mistakenly attribute our own normative outlook to other people.

### 5.3.3. On the prospects of a unified account of guilt

Naturally, the idea that one's normative dominant can have a dramatic influence on the shape of one's guilt feelings raises the question: are there any prospects at all to unite the sphere of guilt under a common definition? In what follows, I will explain the general direction that such a broader account of guilt could take, if it is possible at all, but will also explain why such an account of guilt is likely to become so abstract that, even though it may provide a less judgmental starting point for our inquiry into the terrain of guilt, it might not be of much help in more practical contexts, for example, when addressing people's guilt feelings in therapy. This will go to show that, in addition to working on the unity, it is also important to work on revealing the diversity and individual variance between people's feelings of guilt.

As we have seen, some forms of guilt, especially the ones that arise from a more agent-based deontology, are likely to remain dutifully within the bounds of the traditional, deontic account of guilt. Others, however, like the value-based type of guilt, will bleed out of the boundaries of the deontic because the focus was never on the right but on the good, which is a much broader phenomenon. This raises the question: is it possible to bring the deontic and value-based accounts of guilt under a unified higher-order concept of guilt? While it is not my aim to defend such an account, I would nevertheless like to explain in broad terms what such an account might look like. If there are prospects of a unified account of guilt at all, then I think that a promising solution might come from the work that has been done on the unity of the normative. To the extent that guilt feelings can be both more narrowly deontic and more broadly axiological, the answer to the question about unification is going to depend on the more general question about the unity of the normative sphere. It is common to consider axiological (or evaluative) concepts like the good and the bad, and deontic concepts like the right and the wrong, as belonging to the same broad category of normative concepts, but what exactly is it that might justify this practice? According to Christine Tappolet, the most attractive way of granting unity is by appealing to the idea that both axiological and deontic concepts can be seen as reason-giving (Tappolet, 2014, pp. 51–52). So, the idea is that if something is viewed as good or right, then it is viewed as giving us a reason to act positively towards that thing and, vice versa, if something is bad or wrong, then it gives us a reason to act negatively towards that thing. If we consider the concept of reason as the normative concept *par excellence*, then deontic and axiological concepts would both remain at the same conceptual level, as concepts that can be explained by the idea that they are reason-giving. The main attraction of this approach lies in the fact that it offers a neutral framework that does not prioritise either of those conceptual families, neither the deontic nor the axiological. The alternative would be to attempt to reduce one of those conceptual families to the other, but this would be a controversial move because it would make our account of normativity more favourable towards some moral theories rather than others (saying that the right should be derived from the good would already be decidedly axiological and not

particularly hospitable to certain forms of deontology). But, especially if we want to construct a general theory of guilt that would enable us to explain how guilt can be experienced by people with different moral outlooks, going for a more neutral account is the only viable option.

So, how exactly might we integrate the idea of deontic and evaluative concepts both being reason-giving into a general theory of guilt? For instance, we might say something along the lines: to feel guilt is to perceive oneself as contributing to something for which we have a moral reason not to contribute to. For those who work within an axiological framework, this would translate into the idea that they should not contribute to the morally bad because this is their ultimate source of moral reasons. For an agent-based deontologist, however, this would mean that they should not contribute to the morally wrong, where the latter will narrow the sense of contribution down to the causal exercise of one's agency. In this case, the general definition would be the same, but depending on what the person finds as reason-giving, the specific interpretation of this definition will be subject-relative.

As said, this is not meant as a full-blown defence of a general account of guilt, I am merely pointing towards a direction that a general account of guilt might take if it is to accommodate both deontic and axiological views. But there is nevertheless a lot we can learn from this tentative attempt to formulate a general account of guilt. Making an account more inclusive of radically different normative outlooks will come at a cost: it will make the definition more abstract. Such a definition can still serve as a useful starting point for moral philosophers, who can, depending on the subject's moral outlook, decipher the more specific shape that their guilt is bound to take. However, I doubt that people outside normative ethics could easily operate with such a concept. Without expertise in normative ethics, it is hard to understand what it even means, and I am also somewhat doubtful whether this is the right level of abstraction for other sciences. In particular, I am concerned about the usefulness of such a concept for psychology, especially if the goal is to help people deal with their guilt feelings in the context of therapy. In this context, it would seem more appropriate to work with more specific types of guilt that already capture how people with specific moral outlooks are bound to feel guilt. In other words, it would be important to clarify different guilt-feeling profiles that reflect how a particular person, throughout many situations, is likely to experience guilt. My account of axiological guilt seems to point towards one such profile. Taking this more subject-relative approach is important because, if the aim is to help the subject cope with their guilt feelings (which may also include providing them with the conceptual framework to better make sense of their guilt), our account of guilt must be specific enough to capture what kind of guilt feelings the subject is in fact experiencing. Only then can it capture with precision what the guilt is about, and only then can it point towards solutions with regard to how the subject should deal with their guilt—either by going out there in the world and trying to make it into a better place, or by working on the stained self.

So, perhaps we should not be overly worried if our accounts of guilt do not cover how all people experience guilt. From the viewpoint of grand philosophical aspirations, it may seem like a failure that the value-based account of guilt that I have been building throughout this dissertation cannot serve as an account of all guilt. Similarly, it may also seem like a failure that the moral staining view cannot accommodate all instances of guilt. But, from a practical point of view, those types of accounts may be precisely the accounts that humanity needs the most. At least as long as we steer clear of not proposing them as accounts of all guilt.

## 6. CONCLUSION

The aim of this dissertation was to explicate a wider concept of guilt that emerges from the folk discourse that applies the word ‘guilt’ quite liberally—not just to cases that involve moral wrongdoing but also to those of merely causal harm and survivorship. I hope to have shown that this folk discourse is not just some loose and confused way of talking about guilt but reveals a coherent guilt concept that is unified by the idea that the subject contributes, either causally or constitutively, to a morally bad state of affairs—they make a difference to the world, and unfortunately towards the worse. I call it a value-based account of guilt because the focus is on the moral badness of the states of affairs rather than on the moral wrongness of the agent’s actions. As such, the account offers a much-needed alternative to the traditional, deontic account of guilt, that would deem cases without moral wrongdoing either instances of irrational guilt or just neglect them as something other than guilt. My account enables us to make sense of both guilt for merely causal harm and survivor guilt as meaningful parts of people’s lives.

On the landscape of those theories of guilt that do share the same aim of accommodating a broader range of cases that do not involve moral wrongdoing, my value-based account of guilt stands out for at least three broad reasons: it underscores the moral dimension of the guilt experience, including its tight connection with justice; it integrates survivor guilt as a full-fledged member of the guilt family; and it remains realistically modest in acknowledging its boundaries. Let us review those points in turn.

First, the value-based account of guilt preserves the deep sense in which guilt is a moral emotion. According to the interpersonal account of guilt advanced by social psychologists like Roy Baumeister and his colleagues (1994), guilt is a matter of mere harm to others. So, in order to accommodate cases without moral wrongdoing, the interpersonal account puts the focus on mere harm and, as such, altogether loses the reference to morality. But this involves an important oversight. As I have argued in Chapter 3, the mere sense of having caused harm is not sufficient for guilt because one can feel sincerely sorry about the harm and yet perceive the harm as just—in which case they need not report any sense of guilt. For example, a professor may not feel guilt about giving a bad grade if they perceive the grade as just in every sense. Acknowledging the possibility of such cases helps us see that the cases of guilt that are usually discussed as instances of harm without perceived wrongdoing are, upon closer inspection, cases where the harm is nevertheless seen as unfair towards the victim. For example, when a driver causes a crash by no fault of their own and another human being dies as a result, then their death is nevertheless usually construed as unjust. Thus, guilt is not just a matter of nonmoral harm—the subject must view the harm as morally bad. This is why there is nevertheless an ineliminable moral element involved in guilt. The value-based account of guilt makes this element visible again by shifting the focus from morally wrong actions to morally bad states of affairs. So, even though the subject need not see themselves as having done something morally

wrong, they nevertheless perceive themselves as causing a morally bad state of affairs. More specifically, I have also argued that in cases of mere harm to others, moral badness manifests itself in the form of noncomparative injustice and is predominantly linked to desert—the harm is perceived as undeserved by the victim.

Acknowledging that guilt remains a moral emotion even if it does not involve a reference to moral wrongdoing is important for at least two reasons. First, on the background of general scepticism about overmoralising the shapes of emotions, it is an intriguing claim on its own that there are emotions that have a strictly moral shape—in the sense that their core relational concern involves a reference to a moral concept. Even though moral emotions in this traditional sense of the term may be rare, it would be a mistake to altogether deny their existence. Secondly, it is my hope that explicating this moral dimension is also helpful for those who want to make better sense of their own or their loved one’s guilt feelings—that it helps to clarify what it is that their feeling captures. As I have discussed, for many people who go through a traumatic experience where they have, through no fault of their own, caused fatal harm to another or have survived a catastrophe that others did not, not addressing the moral aspects of their experience has not proven to be particularly helpful. Autobiographical reports attest to the fact that people do not perceive those cases as morally neutral, and feel the need to make sense of “the murky moral area” (Peters, 2023, p. 36). I hope to have shown that moral philosophy can be of help when disentangling where and how the dimension of morality enters into guilt without wrongdoing. It has a rich and nuanced conceptual toolbox that enables us to better differentiate the moral wrongness of actions and the moral wrongness of states of affairs.

Furthermore, when talking about the moral badness of states of affairs, there is also room for further analysis regarding the precise values by which the moral badness manifests itself. As we were able to see, in cases of merely causal harm, matters of noncomparative justice, especially desert, rose to the fore, while survivor guilt was conceptually tied to comparative justice. That cases of guilt that are not covered by deontic norm violations are nevertheless often linked with a broader sense of justice, is an interesting result on its own. Even though contemporary moral philosophers do not usually equate morality with justice (as I explained in 5.1, this would bring problems of its own), there is still something we can learn from a more justice-based approach. Namely, even though in some sense justice is a more narrow concept than the more general talk about moral norms (for example, justice does not seem well suited to capture this part of morality that links with benevolence), there is also another sense in which it can serve as a wider concept. Justice is not exclusively tied to deontic norm violations, but can also be understood in a more general, ‘cosmic’ sense of the term that applies to anything that just happens. In this wider sense, the focus is not so much on how the injustice was brought about but on the idea that, towards the victim, what happened is nevertheless unfair. As such, the discourse of justice has the means to accommodate not just an agent-based view of morality but also a more victim-based view. Furthermore, if the focus is not on what we are morally obliged to do but on what would be an emotionally appropriate response to certain

happenings in the world, then some forms of more victim-based justice like desert would definitely require much closer attention than they have received so far. While I do not think that we should seek to reduce all morality to matters of justice, my value-based account preserves the important sense in which guilt can be linked to the kind of moral badness that is more victim- rather than agent-centred. The value-based account of guilt puts the focus on the moral badness of the states of affairs that one has contributed to, and, very often, this badness is to be analysed in terms of different forms of injustice.

The second main attraction of the type of value-based account of guilt that I have defended is that it enables us to integrate survivor guilt as a full-fledged member of the guilt family. This point is distinct from the previous one, which highlighted the importance of acknowledging the moral dimension of guilt. In the context of survivor guilt, there is a sufficient consensus about guilt being linked to injustice, and I am also confident that if we presented a more detailed typology of injustice, most theorists would agree that the type of injustice that is involved in guilt is a comparative one—it is perceived as unfair that one made it out alive while the others who were no less worthy did not. As such, it is not usually disputed that survivor guilt is a moral emotion. So, highlighting the link with injustice alone adds little new to what has already been agreed upon. The challenge lies in getting injustice into survivor guilt *in the right way*—so that it becomes clear what unifies survivor guilt with other instances of guilt. As I have argued, other alternatives to the traditional account of guilt have struggled with this. In order to grant unification, they end up building unnecessary elements into our conception of survivor guilt—either some ultimately nevertheless causal relation to something bad (Taylor, 1985; Zhao, 2020), the idea that one’s privileged position must come at the expense of other people (Baumeister et al., 1994), or a more radical reconceptualisation of guilt as a matter of anxiety about warranted resentment and about having nothing to say in defence of one’s privilege (Velleman, 2003; MacKenzie & Zhao, 2023). Throughout this dissertation, I have argued that rather than taking us closer to the core of survivor guilt, those approaches only take us further away from the most fundamental aspect of survivor guilt—the very fact that the subject being alive is comparatively unjust towards those who died. My suggestion is that the answer to integrating survivor guilt is simpler than those other accounts make it seem: we only need to recognise that the relationship between the subject and the unjust state of affairs is of constitutive rather than causal kind. Survivor guilt involves perceiving oneself as a constitutive part of the comparatively unjust state of affairs. And the unification is to be found in the idea that both causal and constitutive relations can be captured by a relation of dependence that can be described in terms of difference-making. Regardless of whether the subject causes something morally bad or is a constitutive element in something morally bad, there is a sense in which they nevertheless make a difference towards the worse.

The third advantage of my value-based account of guilt is its modesty. The account does not have a pretension of capturing each and every instance of guilt out there. It captures a unique type of guilt that arises from a value-based moral

outlook where normative priority is given to the goodness of states of affairs, which basically means that the good is considered as prior to the right, so that the rightness of actions is derived from the kinds of states of affairs they bring about. For people who are first and foremost oriented towards the states of affairs that take place out there in the world, guilt is unlikely to remain in the boundaries of the deontic—regardless of whether they have contributed to something bad culpably or through no fault of their own, the most salient feature for them is nevertheless that they have contributed to a morally bad state of affairs, and this focus is likely to reflect itself in a sufficiently similar emotional reaction. I identify this type of guilt based on one fundamental criterion based on which moral theories (and respective moral outlooks) can be differentiated: normative priority. Normative priority reflects what it is that the person first and foremost cares about when they care about morality—is it the goodness of states of affairs, the rightness of actions, virtue of character? Making psychological distinctions based on the nuances of moral theories is a delicate matter because it is not clear that all divisions that are relevant for moral philosophers necessarily translate into a psychologically relevant difference. But I do think that focusing on normative priority is a good anchor point for talking about guilt because it reflects what it is that the person more specifically cares about when they care about morality, and as such it also tracks the focus of attention, which is relevant for the intentional orientation of emotions. Assuming that the normative dominant is a relatively stable element across individual people, differentiating types of guilt based on the normative dominants is also useful for practical purposes—it does not just capture a phenomenon that is scattered here and there; it captures a distinct kind of guilt-feeling profile.

My project can only be considered as the beginning of a bigger project. By analysing the kind of guilt that focuses on the moral badness of states of affairs, I have shown that normative priority can have a dramatic impact on the very shape of guilt feelings. This calls into question whether some previous accounts of guilt, too—like that of Andrew Ingram’s moral staining view—might actually be best described as capturing a kind of guilt that only arises from a more narrowly deontological moral outlook. It also raises the question of whether moral outlooks that focus on virtue as a normative dominant might not result in guilt at all but in shame instead. What exactly was the key difference between ancient honour/shame cultures like Ancient Greece that resulted in a moral psychology that was dominated by shame rather than guilt? Those are all interesting questions for future research.

It draws attention to the idea that we should not impose our own intuitive understanding of guilt as an account of all guilt. This also means that we should not think that if another person reports guilt, we know well what it is that they mean. I started this dissertation by highlighting that oftentimes, people who feel guilt for merely causal harm report being misunderstood. The dissertation ends with an acknowledgement that such misunderstandings may be more common than one might initially think. When we take more extreme cases that remain incomprehensible for the other side, those differences become visible, but since

people's overall moral outlook remains the same across all kinds of guilt scenarios, it should be clear that how they interpret the standard scenario of moral wrongdoing will be different as well—for some, it will still be the badness of the consequences that remains the focal point, whereas for others it may be the idea that they have morally stained themselves.

## BIBLIOGRAPHY

- Alexander, L., & Moore, M. (2024). Deontological Ethics. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Winter 2024). Metaphysics Research Lab, Stanford University.  
<https://plato.sydney.edu.au/entries/ethics-deontological/>
- Aristotle. (1991a). *Nicomachean Ethics* (T. Irwin, Ed. & Trans.). Hackett.
- Aristotle. (1991b). *On Rhetoric* (G. A. Kennedy, Trans.). Oxford University Press.
- Bagnoli, C. (2000). Value in the Guise of Regret. *Philosophical Explorations*, 3(2), 169–187.
- Barrett, L. F. (2017). *How emotions are made: The new science of the mind and brain*. Houghton Mifflin Harcourt.
- Baumeister, R. F., Stillwell, A. M., & Heatherton, T. F. (1994). Guilt: An interpersonal approach. *Psychological Bulletin*, 115(2), 243–267.
- Baumeister, R. F., Stillwell, A. M., & Heatherton, T. F. (1995a). Interpersonal aspects of guilt: Evidence from narrative studies. In J. P. Tangney & K. W. Fischer (Eds.), *Self-conscious emotions: The psychology of shame, guilt, embarrassment, and pride* (pp. 255–273). The Guilford Press.
- Baumeister, R. F., Stillwell, A. M., & Heatherton, T. F. (1995b). Personal Narratives About Guilt: Role in Action Control and Interpersonal Relationships. *Basic and Applied Social Psychology*, 17(1–2), 173–198.
- Baumeister, R. F., & Wotman, S. R. (1994). *Breaking hearts: The two sides of unrequited love*. Guilford Press.
- Bell, M. (2013). *Hard feelings: The moral psychology of contempt*. Oxford University Press.
- Benedict, R. (1946). *The chrysanthemum and the sword: Patterns of Japanese culture*. Houghton Mifflin.
- Bennett, J. (1989). Two departures from consequentialism. *Ethics*, 100(1), 54–66.
- Ben-Ze'ev, A. (2001). *The subtlety of emotions*. MIT Press.
- Ben-Ze'ev, A. (2002). Are envy, anger, and resentment moral emotions? *Philosophical Explorations*, 5(2), 148–154.
- Ben-Ze'ev, A. (2014). The personal comparative concern in schadenfreude. In W. W. van Dijk & J. W. Ouwerkerk (Eds.), *Schadenfreude: Understanding Pleasure at the Misfortune of Others* (pp. 77–90). Cambridge University Press.
- Bernard, V., Ottenberg, P., & Redl, F. (1965). Dehumanization: A composite psychological defense in relation to modern war. *Behavioral Science and Human Survival*, 64–82.
- Bettelheim, B. (1979). *Surviving, and other essays*. Random House Inc.
- Bica, C. M. (1999). A therapeutic application of philosophy. The moral casualties of war: Understanding the experience. *International Journal of Applied Philosophy*, 13(1), 81.
- Bishop, J. C. (1989). *Natural agency: An essay on the causal theory of action*. Cambridge University Press.
- Blacher, R. S. (2000). “It isn’t fair”: Postoperative depression and other manifestations of survivor guilt. *General Hospital Psychiatry*, 22(1), 43–48.
- Blum, L. (1987). Compassion. In R. B. Kruschwitz & R. C. Roberts (Eds.), *The Virtues: Contemporary Essays in Moral Character*. Wadsworth.
- Bradley, B. (2009). *Well-being and death*. Clarendon Press.

- Bradley, B. (2012). Doing away with harm. *Philosophy and Phenomenological Research*, 85(2), 390–412.
- Brand, M. (1984). *Intending and acting: Toward a naturalized action theory*. MIT press Cambridge, MA.
- Bratman, M. (1987). *Intention, plans, and practical reason*. Harvard University Press.
- Brison, S. J. (2002). *Aftermath: Violence and the Remaking of a Self*. Princeton University Press.
- Brock, R. N., & Lettini, G. (2012). *Soul repair: Recovering from moral injury after war*. Beacon Press.
- Brockner, J., Davy, J., & Carter, C. (1985). Layoffs, self-esteem, and survivor guilt: Motivational, affective, and attitudinal consequences. *Organizational Behavior and Human Decision Processes*, 36(2), 229–244.
- Broome, J. (1991). *Weighing goods: Equality, uncertainty and time*. Basil Blackwell.
- Brosnan, S. F., & De Waal, F. B. (2003). Monkeys reject unequal pay. *Nature*, 425(6955), 297–299.
- Brun, G. (2020). Conceptual re-engineering: From explication to reflective equilibrium. *Synthese*, 197(3), 925–954.
- Brun, G. (2022). Re-engineering contested concepts. A reflective equilibrium approach. *Synthese*, 200, 1–29.
- Bush, M. (1989). The role of unconscious guilt in psychopathology and psychotherapy. *Bulletin of the Menninger Clinic*, 53(2), 97–107.
- Byrne, A. (1998). Interpretivism. *European Review of Philosophy*, 3, 199–223.
- Campbell, S. M., Ulrich, C. M., Grady, C., Campbell, S. M., Ulrich, C. M., & Grady, C. (2018). A broader understanding of moral distress. In C. M. Ulrich & C. Grady (Eds.), *Moral distress in the health professions* (pp. 59–77). Springer.
- Carni, S., Petrocchi, N., Del Miglio, C., Mancini, F., & Couyoumdjian, A. (2013). Intrapsychic and interpersonal guilt: A critical review of the recent literature. *Cognitive Processing*, 14(4), 333–346.
- Church, R. M. (1959). Emotional reactions of rats to the pain of others. *Journal of Comparative and Physiological Psychology*, 52(2), 132–134.
- Clarke, R. (2016). Moral responsibility, guilt, and retributivism. *The Journal of Ethics*, 20(1), 121–137.
- Cokelet, B., & Maley, C. J. (Eds.). (2019). *The Moral Psychology of Guilt*. Rowman & Littlefield International.
- Coplan, A. (2011). Will the real empathy please stand up? A case for a narrow conceptualization. *The Southern Journal of Philosophy*, 49, 40–65.
- Corder, C. (2007). Guilt, Remorse and Victims. *Philosophical Investigations*, 30(4), 337–362.
- Cova, F., Deonna, J., & Sander, D. (2015). Introduction: Moral emotions. *Topoi*, 34, 397–400.
- D’Arms, J. (2022). Fitting Emotion. In C. Howard & R. A. Rowland (Eds.), *Fittingness: Essays in the Philosophy of Normativity* (pp. 105–129). Oxford University Press.
- D’Arms, J. (2024). What makes an emotion appropriate or inappropriate to its object? In A. Scarantino (Ed.), *Emotion Theory: The Routledge Comprehensive Guide: Volume II: Theories of Specific Emotions and Major Theoretical Challenges* (pp. 615–637). Taylor & Francis.
- D’Arms, J., & Jacobson, D. (2000). The Moralistic Fallacy: On the ‘Appropriateness’ of Emotions. *Philosophy and Phenomenological Research*, 61(1), 65–90.

- D'Arms, J., & Jacobson, D. (2003). The significance of recalcitrant emotion (or, anti-quasijudgmentalism). *Royal Institute of Philosophy Supplements*, 52, 127–145.
- D'Arms, J., & Jacobson, D. (2022). The Motivational Theory of Guilt (and Responsibility). In A. B. Carlsson (Ed.), *Self-blame and moral responsibility*. Cambridge University Press.
- D'Arms, J., & Jacobson, D. (2023). *Rational sentimentalism*. Oxford University Press.
- Darwall, S. L. (1986). Agent-centered restrictions from the inside out. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 50(3), 291–319.
- Davidson, D. (1980a). Actions, Reasons, and Causes. In *Essays on Actions and Events* (pp. 3–20). Clarendon Press.
- Davidson, D. (1980b). Agency. In *Essays on Actions and Events* (pp. 43–61). Clarendon Press.
- Davidson, D. (2004). *Paradoxes of Irrationality*. Oxford University Press.
- Deigh, J. (1982). Love, guilt, and the sense of justice. *Inquiry*, 25(4), 391–416.
- Deigh, J. (1999). All Kinds of Guilt. *Law and Philosophy*, 18(4), 313–325.
- Deigh, J. (2019). Empathy and Conscience. In C. J. Maley & B. Coker (Eds.), *The Moral Psychology of Guilt* (pp. 53–70). Rowman & Littlefield International.
- Dennett, D. C. (1987). *The Intentional Stance*. MIT Press.
- Dow, P. (2000). *Physical Causation*. Cambridge University Press.
- Dretske, F. (1986). Misrepresentation. In R. Brogdan (Ed.), *Belief: Form, content and function* (pp. 17–36). Oxford University Press.
- Dretske, F. (1988). *Explaining behavior: Reasons in a world of causes*. MIT Press.
- Driver, J. (1996). The virtues and human nature. In R. Crisp (Ed.), *How should one live?: Essays on the virtues* (pp. 111–129). Clarendon Press.
- Eger, E. E. (2017). *The choice: Embrace the possible*. Simon and Schuster.
- Elison, J. (2005). Shame and guilt: A hundred years of apples and oranges. *New Ideas in Psychology*, 23(1), 5–32.
- Enç, B. (2003). *How we act: Causes, reasons, and intentions*. Oxford University Press.
- Feinberg, J. (1970). *Doing & deserving: Essays in the theory of responsibility*. Princeton University Press.
- Feinberg, J. (1974). Noncomparative Justice. *Philosophical Review*, 83, 297–338.
- Feinberg, J. (1984). *Harm to others*. Oxford University Press.
- Fischer, J. (2017). Pride and moral responsibility. *Ratio*, 30(2), 181–196.
- Foot, P. (1969). Moral Beliefs. In W. D. Hudson (Ed.), *Is-Ought Question*. Springer.
- Freud, S. (1962). *Civilization and Its Discontents* (J. Strachey, Trans.). W. W. Norton & Company Inc.
- Friedman, M. (1985). Survivor guilt in the pathogenesis of anorexia nervosa. *Psychiatry*, 48(1), 25–39.
- Fukuma, S. (2013). Rethinking survivor guilt: An attempt at a philosophical interpretation. In *Applied Ethics: Risk Justice and Liberty* (pp. 52–65). Center for Applied Ethics and Philosophy.
- Gangemi, A., & Mancini, F. (2011). Guilt and guilts. *Re-Constructing Emotional Spaces: From Experience to Regulation*, 169–188.
- Gardner, J. (2018). *From personal life to private law*. Oxford University Press.
- Garger, K., Pagonis, S., & Jaeger, M. (2019, January 7). Man whom firefighter died trying to save feels 'guilty'. *New York Post*. <https://nypost.com/2019/01/07/man-who-firefighter-died-trying-to-save-feels-guilty/>

- Gaus, G. F. (2001). What is deontology? Part two: Reasons to act. *Journal of Value Inquiry*, 35, 179–193.
- Gert, B., & Gert, J. (2020). The definition of morality. In *The Stanford encyclopedia of philosophy*. <https://plato.stanford.edu/entries/morality-definition/>
- Gibbard, A. (1990). *Wise choices, apt feelings: A theory of normative judgment*. Harvard University Press.
- Gilbert, M. (1997). Group wrongs and guilt feelings. *The Journal of Ethics*, 1, 65–84.
- Goldman, A. I. (1970). *Theory of human action*. Prentice-Hall.
- Gordon-Solmon, K. (2015). Can Comparative Desert do Without Equality? *Philosophical Papers*, 44(2), 189–205.
- Gordon-Solmon, K. (2017). Comparative Desert Vs. Fairness. *Law and Philosophy*, 36, 367–387.
- Gray, M. (2021). Unintentional killing: A neglected trauma. *Journal of Traumatic Stress Disorders & Treatment*, 10(5), 1–8.
- Gray, M. J., & Yaw, C. (2024). *Accidental Killing: A Survivor's Handbook*. Independently published.
- Greenspan, P. S. (1992). Subjective guilt and responsibility. *Mind*, 101(402), 287–303.
- Greenspan, P. S. (1994). Guilt and virtue. *The Journal of Philosophy*, 91(2), 57–70.
- Greenspan, P. S. (1995). *Practical guilt: Moral dilemmas, emotions, and social norms*. Oxford University Press on Demand.
- Gregory, A. (2017). The Sorrow and Shame of the Accidental Killer. *The New Yorker*. <https://www.newyorker.com/magazine/2017/09/18/the-sorrow-and-the-shame-of-the-accidental-killer>
- Griffin, J. (1986). *Well-Being: Its Meaning, Measurement and Moral Importance*. Clarendon Press.
- Griffioen, A. (2014). Regaining the ‘lost self’: A philosophical analysis of survivor’s guilt. In A. Gerner & J. Golçalves (Eds.), *Altered self and self experience* (pp. 43–57). Norderstedt: Journals on Demand.
- Griffiths, P. E. (1997). *What emotions really are: The problem of psychological categories*. University of Chicago Press.
- Guralnick, P. (1994). *Last train to Memphis: The rise of Elvis Presley*. Little, Brown and Co.
- Haidt, J. (2003). The moral emotions. *Handbook of Affective Sciences*, 11(2003), 852–870.
- Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. Pantheon Books.
- Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, 133(4), 55–66.
- Hammerton, M. (2025). The fundamental divisions in ethics. *Inquiry*, 68(2), 318–341.
- Harcourt, E. (2013). Guilt, shame, and the ‘psychology of love’. In L. Braddock & M. Lacey (Eds.), *The Academic Face of Psychoanalysis* (pp. 133–147). Routledge.
- Harcourt, E. (2023). Shame, Guilt and the ‘Morality System’. *Passion: Journal of the European Philosophical Society for the Study of Emotions*, 1(2), 125–138.
- Hare, R. M. (1952). *The language of morals*. Oxford University Press.
- Harpham, W. S. (2018). A Better Name for Survivor Guilt. *Oncology Times*, 40(9), 18.
- Harris, G. W. (1997). *Dignity and vulnerability: Strength and quality of character*. University of California Press.
- Harris, J. (1975). The Survival Lottery. *Philosophy*, 50(191), 81–87.

- Hoffman, M. L. (2000). *Empathy and moral development: Implications for caring and justice*. Cambridge University Press.
- Howard-Snyder, F. (1994). The heart of consequentialism. *Philosophical Studies*, 76, 107–129.
- Hume, D. (1974). *An Enquiry Concerning Human Understanding*. Doubleday.
- Hurka, T. (1992). Consequentialism and content. *American Philosophical Quarterly*, 29(1), 71–78.
- Hurka, T. (2000). *Virtue, vice, and value*. Oxford University Press.
- Hurka, T. (2001). The common structure of virtue and desert. *Ethics*, 112(1), 6–31.
- Hutto, D. D. (2012). Truly enactive emotion. *Emotion Review*, 4(2), 176–181.
- Ingram, A. (2017). Guilt, Practical Identity, and Moral Staining. *Philosophy*, 92(4), 623–645.
- Ingram, A. (2020). Guilt feelings and the intelligibility of moral duties. *Ratio*, 33(1), 56–67.
- Jäger, C., & Bartsch, A. (2006). Meta-Emotions. *Grazer Philosophische Studien*, 73(1), 179–204.
- Juni, S. (2016). Survivor guilt: A critical review from the lens of the Holocaust. *International Review of Victimology*, 22(3), 321–337.
- Kagan, S. (1998). *Normative ethics*. Westview.
- Kagan, S. (2003). Comparative desert. In Olsaretti (Ed.), *Desert and justice* (pp. 93–122). Oxford University Press.
- Kagan, S. (2014). *The geometry of desert*. Oxford University Press.
- Kagan, S. (2017). Equality and desert. In A. Mancilla & T. Campbell (Eds.), *Theories of justice* (pp. 439–456). Routledge.
- Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–208). Cambridge University Press.
- Kekes, J. (1983). Constancy and purity. *Mind*, 92(368), 499–518.
- Kenny, A. (1963). *Action, Emotion, and Will*. Routledge.
- Kinghorn, K. (2021). *The Nature of Desert Claims: Rethinking what it Means to Get One's Due*. Cambridge University Press.
- Kohlberg, L. (1981). *The philosophy of moral development: Moral stages and the idea of justice*.
- Korsgaard, C. M. (1996). *The sources of normativity*. Cambridge University Press.
- Korsgaard, C. M. (2009). *Self-constitution: Agency, identity, and integrity*. Oxford University Press.
- Kriegel, U. (2015a). The character of cognitive phenomenology. In T. Breyer & C. Gutland (Eds.), *Phenomenology of Thinking* (pp. 25–43). Routledge.
- Kriegel, U. (2015b). *The varieties of consciousness*. Oxford University Press.
- Kristjánsson, K. (2006). *Justice and desert-based emotions*. Aldershot.
- Kristjánsson, K. (2014). Is shame an ugly emotion? Four discourses—Two contrasting interpretations for moral education. *Studies in Philosophy and Education*, 33(5), 495–511.
- Kurth, C. (2018). *The anxious mind: An investigation into the varieties and virtues of anxiety*. Mit Press.
- Kurth, C. (forthcoming). When Is an Emotion an Adaption? A Look at Shame and Guilt. *Passion: Journal of the European Philosophical Society for the Study of Emotions*, 3(1).
- La Caze, M. (2002). Revaluating envy and resentment. *Philosophical Explorations*, 5(2), 155–158.

- Lamb, R. (1983). Guilt, shame, and morality. *Philosophy and Phenomenological Research*, 43(3), 329–346.
- Landman, J. (1987). Regret: A theoretical and conceptual analysis. *Journal for the Theory of Social Behaviour*, 132–160.
- Lazarus, R. S. (1984). The trivialization of distress. In B. L. Hammonds & C. J. Scheier (Eds.), *Psychology and Health: The master lecture series* (pp. 121–144). American Psychological Association.
- Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford University Press.
- Lerner, M. J., & Matthews, G. (1967). Reactions to suffering of others under conditions of indirect responsibility. *Journal of Personality and Social Psychology*, 5(3), 319–325.
- Levi, P. (1988). *The drowned and the saved* (R. Rosenthal, Trans.). Michael Joseph.
- Lewis, D. (2000). Causation as influence. *The Journal of Philosophy*, 97(4), 182–197.
- Lewis, H. B. (1971). *Shame and guilt in neurosis*. International Universities Press.
- Leys, R. (2007). *From Guilt to Shame: Auschwitz and After*. Princeton University Press.
- Lifton, R. J. (1991). *Death in life: Survivors of Hiroshima*. University of North Carolina Press.
- Lifton, R. J. (1996). *The broken connection: On death and the continuity of life*. American Psychiatric Pub.
- Lindsay-Hartz, J. (1984). Contrasting experiences of shame and guilt. *American Behavioral Scientist*, 27(6), 689–704.
- Lindsay-Hartz, J., De Rivera, J., & Mascolo, M. F. (1995). Differentiating guilt and shame and their effects on motivation. In J. P. Tangney & K. Fischer (Eds.), *Self-conscious emotions: The psychology of shame, guilt, embarrassment, and pride* (pp. 274–300). Guilford Press.
- Linchan, A. (2019, November 11). I Watched Friends Die in Afghanistan. The Guilt Has Nearly Killed Me. *The New York Times*.  
<https://www.nytimes.com/2019/11/11/magazine/survivor-guilt-veteran.html>
- Lupfer, M. B., & Gingrich, B. E. (1999). When bad (good) things happen to good (bad) people: The impact of character appraisal and perceived controllability on judgments of deservingness. *Social Justice Research*, 12, 165–188.
- MacIntyre, A. (1981). *After virtue*. Duckworth.
- MacKenzie, J., & Zhao, M. (2023). Survivor guilt. *Philosophical Studies*, 180(9), 2707–2726.
- Maibom, H. L. (2019). On the Distinction between Shame and Guilt. In C. Bradford & C. J. Maley (Eds.), *The Moral Psychology of Guilt* (pp. 37–51). Rowman & Littlefield International.
- Maley, C. J. (2015). *On the Nature of Guilt and Shame*. Princeton University.
- Maley, C. J., Harman, G., Maley, C. J., & Cokelet, B. (2019). The Feeling of Guilt. In *The Moral Psychology of Guilt* (pp. 13–36). Rowman & Littlefield International.
- Mancini, F., & Gangemi, A. (2021). Deontological and altruistic guilt feelings: A dualistic thesis. *Frontiers in Psychology*, 12, 1–16.
- Mandel, D. (2003). Counterfactuals, emotions, and context. *Cognition and Emotion*, 17(1), 139–159.
- Masserman, J. H., Wechkin, S., & Terris, W. (1964). ‘Altruistic’ behavior in rhesus monkeys. *American Journal of Psychiatry*, 121(6), 584–585.
- Matsakis, A. (1999). *Survivor guilt: A self-help guide*. New Harbinger Publications.
- McClay, W. M. (2017). The strange persistence of guilt. *The Hedgehog Review*, 19(1), 40–55.

- McFall, L. (1987). Integrity. *Ethics*, 98(1), 5–20.
- McNaughton, D., & Rawling, P. (1991). Agent-relativity and the doing-happening distinction. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 63(2), 167–185.
- McNaughton, D., & Rawling, P. (1992). Honoring and promoting values. *Ethics*, 102(4), 835–843.
- McNaughton, D., & Rawling, P. (1998). On defending deontology. *Ratio*, 11(1), 37–54.
- Mele, A. R. (1992). *Springs of action: Understanding intentional behavior*. Oxford University Press.
- Mele, A. R. (2003). *Motivation and agency*. Oxford University Press.
- Metz, T. (2018). Making sense of survivor's guilt: Why it is justified by an African ethic. In G. Hull (Ed.), *Debating African Philosophy* (pp. 148–163). Routledge.
- Miceli, M., & Castelfranchi, C. (1998). How to silence one's conscience: Cognitive defenses against the feeling of guilt. *Journal for the Theory of Social Behaviour*, 28(3), 287–318.
- Miceli, M., & Castelfranchi, C. (2018). Reconsidering the differences between shame and guilt. *Europe's Journal of Psychology*, 14(3), 710–733.
- Mill, J. S. (1998). *Utilitarianism* (R. Crisp, Ed.). Oxford University Press.
- Miller, D. (1976). *Social Justice*. Clarendon Press.
- Miller, D. (1999). *Principles of social justice*. Harvard University Press.
- Miller, S. C. (2015). The Moral Meanings of Miscarriage. *Journal of Social Philosophy*, 46(1), 141–157.
- Miyazono, K., & Bortolotti, L. (2021). *Philosophy of psychology: An introduction*. John Wiley & Sons.
- Modell, H., Arnold. (1965). On having the right to a life: An aspect of the superego's development. *International Journal of Psycho-Analysis*, 46, 323–331.
- Modell, H., Arnold. (1971). The origin of certain forms of pre-oedipal guilt and the implications for a psychoanalytic theory of affects. *International Journal of Psycho-Analysis*, 52, 337–346.
- Mölder, B. (2010). *Mind Ascribed: An Elaboration and Defence of Interpretivism*. John Benjamins.
- Morris, H. (1976). *On guilt and innocence: Essays in legal philosophy and moral psychology*. University of California Press.
- Morris, H. (1987). Nonmoral Guilt. In F. Shoeman (Ed.), *Responsibility, character, and the emotions: New essays in moral psychology*. Cambridge University Press.
- Morris, H. (1999). Some further reflections on guilt and punishment. *Law and Philosophy*, 18, 363–378.
- Morton, A. (2013). Imaginary emotions. *The Monist*, 96(4), 505–516.
- Mulligan, K. (2008). Moral Emotions. In D. Sander & K. R. Scherer (Eds.), *Oxford companion to the affective sciences*. Oxford University Press.
- Murphy, J. G. (2012). Shame Creeps Through Guilt and Feels Like Retribution. In *Punishment and the moral emotions: Essays in law, morality, and religion* (pp. 94–113). Oxford University Press.
- Nagai, T. (1951). *We of Nagasaki: The Story of Survivors in an Atomic Wasteland* (I. Shirato & H. B. Silverman, Trans.). Duell, Sloan and Pearce.
- Ney, A. (2014). *Metaphysics: An introduction*. Routledge.
- Niederland, W. G. (1961). The problem of the survivor. *Journal of the Hillside Hospital*, 10(3–4), 233–247.

- Nietzsche, F. (2007). *On the Genealogy of Morality* (K. Ansell-Pearson, Ed.; C. Diethe, Trans.). Cambridge University Press.
- Norcross, A. (2005). Harming in context. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 123(1/2), 149–173.
- Nussbaum, M. (1996). Compassion: The basic social emotion. *Social Philosophy and Policy*, 13(1), 27–58.
- Nussbaum, M. C. (2001). *The fragility of goodness: Luck and ethics in Greek tragedy and philosophy*. Cambridge University Press.
- Nussbaum, M. C. (2003). *Upheavals of thought: The intelligence of emotions*. Cambridge University Press.
- Nussbaum, M. C. (2016). *Anger and forgiveness: Resentment, generosity, justice*. Oxford University Press.
- O'Connor, L. E., Berry, J. W., Weiss, J., Bush, M., & Sampson, H. (1997). Interpersonal guilt: The development of a new measure. *Journal of Clinical Psychology*, 53(1), 73–89.
- Olsaretti, S. (2003). Introduction: Debating desert and justice. In S. Olsaretti (Ed.), *Desert and Justice* (pp. 1–24). Clarendon Press.
- O'Neill, O. (1996). *Towards justice and virtue: A constructive account of practical reasoning*. Cambridge University Press.
- Ortony, A. (1987). Is guilt an emotion? *Cognition and Emotion*, 1(3), 283–298.
- Parfit, D. (1984). *Reasons and Persons*. Oxford University Press.
- Peters, D. W. (2023). *Accidental: Rebuilding a life after taking one*. Broadleaf Books.
- Pethania, Y., Murray, H., & Brown, D. (2018). Living a life that should not be lived: A qualitative analysis of the experience of survivor guilt. *Journal of Traumatic Stress Disorders and Treatment*, 7(1), 1–6.
- Pettit, P. (1989). Consequentialism and respect for persons. *Ethics*, 100(1), 116–126.
- Pettit, P. (1997). The consequentialist perspective. In M. W. Baron, P. Pettit, & M. Slote (Eds.), *Three methods of ethics* (pp. 92–174). Oxford: Blackwell.
- Pojman, L. (2001). Justice as desert. *Queensland University of Technology Law & Justice Journal*, 16, 88–109.
- Pojman, L., & Fieser, J. (2012). *Ethics: Discovering Right and Wrong* (7th ed.). Wadsworth, Cengage Learning.
- Portmann, J. (2000). *When bad things happen to other people*. Routledge.
- Portmore, D. W. (2001). Can an act-consequentialist theory be agent-relative? *American Philosophical Quarterly*, 38(4), 363–377.
- Portmore, D. W. (2003). Position-relative consequentialism, agent-centered options, and supererogation. *Ethics*, 113(2), 303–332.
- Portmore, D. W. (2011). *Commonsense consequentialism: Wherein morality meets rationality*. Oxford University Press.
- Prinz, J. (2004). *Gut reactions: A perceptual theory of emotion*. Oxford University Press.
- Prinz, J. (2008). Is morality innate? In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Vol. 1: The Evolution of Morality: Adaptions and Innateness* (pp. 367–406). MIT Press.
- Prinz, J. (2009). The Moral Emotions. In P. Goldie (Ed.), *The Oxford handbook of philosophy of emotion* (pp. 519–538). Oxford University Press.
- Prinz, J. (2011a). Against empathy. *The Southern Journal of Philosophy*, 49, 214–233.
- Prinz, J. (2011b). Is empathy necessary for morality? In A. Coplan & P. Goldie (Eds.), *Empathy: Philosophical and psychological perspectives* (pp. 211–229). Oxford University Press.

- Prinz, J., & Nichols, S. B. (2010). Moral emotions. In J. M. Doris (Ed.), *The moral psychology handbook* (pp. 519–538). Oxford University Press.
- Ramsey, G., & Deem, M. J. (2022). Empathy and the Evolutionary Emergence of Guilt. *Philosophy of Science*, 89(3), 434–453.
- Rawls, J. (2005). *A theory of justice*. Harvard University Press.
- Roberts, R. C. (2003). *Emotions: An essay in aid of moral psychology*. Cambridge University Press.
- Roberts, R. C. (2009). The vice of pride. *Faith and Philosophy*, 26(2), 119–133.
- Roese, N. J. (2001). The crossroads of affect and cognition: Counterfactuals as compensatory cognitions. In G. Moskowitz (Ed.), *Cognitive social psychology: The Princeton Symposium on the legacy and future of social cognition* (pp. 307–316). Lawrence Erlbaum Mahwah, NJ.
- Ross, W. D. (2002). *The right and the good*. Oxford University Press.
- Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology*, 76(4), 574–586.
- Sadler, B. J. (2016). What’s Love Got to Do with It? Morality and Personal Relationships. In E. E. Kroeker & K. Schaubroeck (Eds.), *Love, Reason and Morality* (pp. 125–144). Routledge.
- Sadurski, W. (1985). *Giving desert its due: Social justice and legal theory*. Springer Science & Business Media.
- Sartorio, C. (2005). Causes as difference-makers. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 123(1/2), 71–96.
- Scarantino, A. (2012). How to define emotions scientifically. *Emotion Review*, 4(4), 358–368.
- Scarantino, A. (2014). The motivational theory of emotions. In D. Jacobson & J. D’Arms (Eds.), *Moral psychology and human agency* (pp. 156–185). Oxford University Press.
- Scarantino, A., & de Sousa, R. (2018). Emotion. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2018). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/win2018/entries/emotion/>
- Schaffer, J. (2004). Causes need not be physically connected to their effects: The case for negative causation. In C. Hitchcock (Ed.), *Contemporary debates in philosophy of science* (pp. 197–216). Blackwell Oxford.
- Scheler, M. (1987). Repentance and rebirth. In *Person and Self-Value: Three Essays* (pp. 87–124). Springer.
- Schlosser, M. (2019). Agency. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2019). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/win2019/entries/agency/>
- Sen, A. (1982). Rights and Agency, *Philosophy and Public Affairs*, 11(1), 3–39.
- Shay, J. (2010). *Achilles in Vietnam: Combat trauma and the undoing of character*. Simon and Schuster.
- Shay, J. (2014). Moral injury. *Psychoanalytic Psychology*, 31(2), 182–191.
- Sherman, N. (2013). Guilt in war. In J. Deigh (Ed.), *On emotions: Philosophical essays* (pp. 179–197). Oxford University Press.
- Shweder, R. A., Much, N. C., Mahapatra, M., & Park, L. (1997). The “big three” of morality (autonomy, community, divinity) and the “big three” explanations of suffering. In A. M. Brandt & P. Rozin (Eds.), *Morality and health* (pp. 119–169). Routledge.

- Sinnott-Armstrong, W. (2005). You ought to be ashamed of yourself (when you violate an imperfect moral obligation). *Philosophical Issues*, 15, 193–208.
- Sinnott-Armstrong, W. (2023). Consequentialism. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Winter 2023). Metaphysics Research Lab, Stanford University.  
<https://plato.stanford.edu/archives/win2023/entries/consequentialism/>
- Skorupski, J. (1995). Agent-neutrality, consequentialism, utilitarianism... a terminological note. *Utilitas*, 7(1), 49–54.
- Sludds, K. (2009). *Emotions: Their cognitive base and ontological importance*. Peter Lang.
- Solomon, R. C. (2003). *Not passion's slave: Emotions and choice*. Oxford University Press.
- Solomon, R. C. (2008). *True to our feelings: What our emotions are really telling us*. Oxford University Press.
- Steinbock, A. J. (2014). *Moral emotions: Reclaiming the evidence of the heart*. Northwestern University Press.
- Stich, S., Doris, J. M., & Roedder, E. (2010). Altruism. In J. M. Doris & the Moral Psychology Research Group (Eds.), *The Moral Psychology Handbook* (pp. 147–205). Oxford University Press.
- Stocker, M. (1973). Rightness and goodness: Is there a difference? *American Philosophical Quarterly*, 10(2), 87–98.
- Stratton-Lake, P. (2002). Introduction. In W. D. Ross, *The right and the good* (pp. ix–1). Oxford University Press.
- Sutrop, M. (2020). Deep Conceptual Moral Disagreements: Over What do We Disagree and Why? *Trames*, 24(3), 295–314.
- Swanton, C. (1995). Profiles of the Virtues. *Pacific Philosophical Quarterly*, 76(1), 47–72.
- Swanton, C. (1997). Virtue ethics and the problem of indirection: A pluralistic value-centred approach. *Utilitas*, 9(2), 167–181.
- Tangney, J. P., & Dearing, R. L. (2003). *Shame and guilt*. Guilford Press.
- Tangney, J. P., Miller, R. S., Flicker, L., & Barlow, D. H. (1996). Are shame, guilt, and embarrassment distinct emotions? *Journal of Personality and Social Psychology*, 70(6), 1256–1269.
- Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral emotions and moral behavior. *Annual Review of Psychology*, 58(1), 345–372.
- Tappolet, C. (2013). Evaluative vs. Deontic concepts. *International Encyclopedia of Ethics*.
- Tappolet, C. (2014). The normativity of evaluative concepts. In A. Reboul (Ed.), *Mind, Values, and Metaphysics: Philosophical Essays in Honor of Kevin Mulligan, Vol II* (pp. 39–54). Springer.
- Taylor, G. (1985). *Pride, shame, and guilt: Emotions of self-assessment*.
- Teroni, F., & Bruun, O. (2011). Shame, Guilt and Morality. *Journal of Moral Philosophy*, 8(2), 223–245.
- Teroni, F., & Deonna, J. A. (2008). Differentiating shame from guilt. *Consciousness and Cognition*, 17(3), 725–740.
- Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge University Press.
- Turner Hospital, J. (2003). *North of Nowhere, South of Loss*. University of Queensland Press.
- Vallentyne, P. (1987). The teleological/deontological distinction. *The Journal of Value Inquiry*, 21, 21–32.

- Velleman, J. D. (2003). XIV. Don't Worry, Feel Guilty\*. *Royal Institute of Philosophy Supplements*, 52, 235–248.
- Warneken, F., & Tomasello, M. (2009). The roots of human altruism. *British Journal of Psychology*, 100(3), 455–471.
- Watanabe, S., & Ono, K. (1986). An experimental analysis of “empathic” response: Effects of pain reactions of pigeon upon other pigeon’s operant behavior. *Behavioural Processes*, 13(3), 269–277.
- Wayment, H. A., Silver, R. C., & Kemeny, M. E. (1995). Spared at Random: Survivor Reactions in the Gay Community 1. *Journal of Applied Social Psychology*, 25(3), 187–209.
- Weiss, J. (1986). Unconscious guilt. *The Psychoanalytic Process: Theory, Clinical Observation and Empirical Research*, 43–67.
- Wiesel, E. (2006). *Night*. Hill and Wang. A division of Farrar, Straus and Giroux.
- Williams, B. (1993). *Shame and necessity*. Taylor & Francis.
- Williams, B. (2006). *Ethics and the Limits of Philosophy*. Routledge.
- Williams, B. A. (1976). Moral Luck. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 50, 115–135.
- Williams, T. (1988). Diagnosis and Treatment of Survivor Guilt. In J. P. Wilson, Z. Harel, & B. Kahana (Eds.), *Human Adaptation to Extreme Stress: From the Holocaust to Vietnam* (pp. 319–336). Springer US.
- Wojtowicz, J. (2019). *Agent-Regret in Our Lives* [PhD Thesis]. King’s College London.
- Wojtowicz, J. (2022). The Purity of Agent-Regret. *Philosophy*, 97(1), 71–90.
- Wolf, S. (2001). The Moral of Moral Luck. *Philosophic Exchange*, 31(1), 5–19.
- Wong, D. B. (2013). Morality, Definition of. In H. LaFollette (Ed.), *The international encyclopedia of ethics* (pp. 3433–3438). Blackwell Publishing Ltd.
- Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford University Press.
- Ylikoski, P. (2013). Causal and constitutive explanation compared. *Erkenntnis*, 78, 277–297.
- Zeelenberg, M., & Van Dijk, E. (2007). On the comparative nature of regret. In D. R. Mandel, D. J. Hilton, & P. Catellani (Eds.), *The psychology of counterfactual thinking* (pp. 159–174). Routledge.
- Zhao, M. (2020). Guilt without perceived wrongdoing. *Philosophy & Public Affairs*, 48(3), 285–314.

## SUMMARY IN ESTONIAN

### Süütunne: kaasates juhtumeid, kus pole valesti toimitud

Tavaliselt räägitakse süütundest kui emotsioonist, mis seondub sellega, et ollakse moraalselt eksinud ja valesti toiminud. Samas raporteeritakse süütunnet tihti ka olukordades, kus ei tajuta end tingimata eksinuna. Rõhutaksin siinkohal eeskätt kahte tüüpjuhtumit: pelgalt kausaalset süüd ja ellujääja süütunnet. Pelgalt kausaalse süü puhul ollakse küll teistele kahju põhjustanud, kuid süütul viisil, ilma et oleks ise midagi valesti tehtud. Seda võib tunda näiteks autoõnnetuste puhul, mis võtsid kellegi elu, kuid mille puhul juht poleks saanud midagi teha, et õnnetust ära hoida (Gregory, 2017; Peters, 2023), aga ka olukordades, kus murtakse kellegi süda, ilma et seejuures usutaks, et meil oleks moraalselt kohustust teisele inimese armastust jagada (Baumeister et al., 1995a). Teiseks on juhtumid, kus inimene tajub, et tal on teistega võrreldes ebaõiglane privileeg. Need juhtumid on tuntud eeskätt ellujääja süü näitel, mida tuntakse pärast inimtekkelisi või looduskatastroofe, kus subjekt jäi ellu, aga teised, kes ei olnud kuidagi halvemad kui tema, hukkusid. Pelgalt kausaalse süü ja ellujääja süü juhtumid esitavad väljakutse traditsioonilisele süükäsitlusele, mis nõuab seost moraalse eksimusega. Käesoleva doktoritöö eesmärgiks on pakkuda alternatiivi traditsioonilisele käsitlusele, mis võimaldaks selgitada seda, miks süütundest võib rääkida ka laiemalt, mitte üksnes moraalse eksimuse juhtudel.

Doktoritöö esimeses, sissejuhatavas peatükis, selgitan lähemalt oma põhiküsimust ning selle tähtsust nii praktikas – inimeste elude kontekstis – kui ka süütunde teemalisel teooriamaastikul. Lähtun töös rahvapsühholoogilises diskursuses levinud avaramast süümõistest, sest leian, et see on oluline, et filosoofilised süüteooriad kõnetaksid seda, mida inimesed ka ise süütunde all mõistavad. Küll aga soovin rõhutada, et kõige sügavamal tasandil seisneb käesoleva doktoritöö põhiküsimus siiski selles, mis see täpsemalt on, mille alusel võiks õigustada rahvapsühholoogilist praktikat vaadelda moraalse eksimuse, pelgalt kausaalse vastutuse ja ellujääja süü juhtumeid samasse kategooriasse kuuluvana. Niisiis ei ole esmatähtis mitte vaidlus sõnade üle (kas seda üldist kategooriat peaks tingimata 'süüks' või millekski muuks nimetama?), vaid see, millises suhtes antud nähtused omavahel on. Avaram süümõiste on praktikas oluline sellepärast, et see aitab meil inimesi paremini mõista. Neil, kes ise sellist süüd tunnevad, võiks avaram süümõiste aidata paremini oma emotsioone lahti mõtestada. Kuid see võiks aidata ka teistel paremini aru saada, milles nende emotsioon seisneb ning kuidas nende suhtes tundlikum olla. Näiteks raporteerivad mitmed avaramat süütunnet tundvad inimesed frustratsiooni, kui oma süütundest teiste inimestega rääkida soovides saadakse vastuseid stiilis „Aga sa ei teinud ju midagi valesti“, mis kipuvad mõjuma pigem vestluselõpetajate kui -algatajatena, sest need ei taba seda, milles antud inimese süütunne tegelikult seisneb (Gregory, 2017; Peters, 2023). Avarama süükäsitluse otsingud on olulised ka teoreetilises plaanis – levinumate süükäsitluste maastikul. Üldiselt on filosoofias levinud pigem traditsiooniline

süükäsitlus, mille kohaselt süütunne seondub ebamoraalse käitumisega, ning eksimusega süü juhtumid on käsitlust leidnud pigem vähe. Selles suhtes on filosoofia olnud samm maas näiteks võrreldes sotsiaalpsühholoogiaga, kus juba ammu on interpersonaalse süükäsitluse all rõhutatud, et fookus tuleks panna pigem teistele kahju tegemisele kui moraalsele eksimusele (Baumeister et al., 1994). Minu töö lähtub samast stardipositsioonist, millest ka interpersonaalne süükäsitlus – tõdemusest, et kui inimesed raporteerivad süüd rohkemates olukordades, kui vaid neis, kus peetakse end moraalselt eksinuks, siis peaks ka meie süükäsitlus sellega arvestama, mitte neid juhtumeid lihtsalt kõrvale lükkama. Küll aga soovin avarama süükäsitluse teematist arutelu moraalifilosoofina edasi arendada, tuues süükäsitluse tagasi moraalse dimensiooni, mis interpersonaalses käsitluses kipub kaduma minema.

Teises peatükis selgitan põhjalikumalt üle, miks traditsiooniline süükäsitlus, mille kohaselt süütunne seisneb valusas tajumuses, et ollakse moraalselt eksinud, ei paku meile abistavat raamistikku pelgalt kausaalse süü ja ellujääja süü mõistmiseks – mis omakorda motiveerib kogu järgnevat tööd kui parema alternatiivi otsimist. Laias laastus on traditsioonilise süükäsitluse pooldajatel kaks strateegiat eksimusega süü juhtumitele vastamiseks. Esiteks võidakse osutada, et inimene peab siiski kasvõi irratsionaalselt uskuma, et on valesti toiminud. Pole küll välisatunud, et mõnikord võib see ka nii olla, aga need juhtumid, millest käesolevas töös lähtun, on juba puhastatud juhtumid, kus subjekt tõesti ei paista arvavat, et on ise midagi valesti teinud, ning nende puhul tuleks heauskusprintsipi kohaselt siiski lähtuda subjekti enda tõlgendusest, kui pole muud tõsist alust neile irratsionaalset uskumust omistada (irratsionaalse uskumuse projitseerimine kellelegi pelgalt sellepärast, et see võimaldaks traditsioonilise süükäsitluse pooldajal oma teooriat kaitsta, ei saa siinkohal pidada tõsiseks aluseks, sest see ei lähtu kõne all olevast fenomenist endast). Teine võimalik strateegia on väita, et tegu pole siiski süütundega. Kas või mida peaks või ei peaks nimetama 'süütundeks' viib keerulisse metateoreetilisse vaidlusse, sest erinevates kontekstides võivad toimida erinevad mõistelised raamistikud. Rõhutan siiski, et praktikas on süüsiildi eitamine viinud tavaliselt selleni, et avaramale süütundele vastav emotsioon lükatakse ebahuvitava uurimisobjektina kõrvale. Lisaks räägib avarama süükäsitluse kasuks ka asjaolu, et vastav rahvapsühholoogiline diskursus on juba olemas. Kui tahame, et meie filosoofilised süükäsitlused kõnetaksid seda, mida inimesed üldiselt süütunde all mõistavad, siis tuleks ka sellise avarama mõistekasutusega arvestada.

Kolmas ja neljas peatükk keskenduvad sellele, kuidas mõista pelgalt kausaalse süü ja ellujääja süü juhtumeid. Kui nende puhul pole keskne see, et ollakse moraalselt eksinud, siis mis see on, mis teeb neist süü – mis see on, mis on keskne?

Kolmandas peatükis on fookuses kausaalse süü juhtumid. Eeskätt just nendest juhtumitest tõukudes arendati välja interpersonaalne süüteooria (Baumeister et al., 1994), mille kohaselt süütunne seisneb kahju põhjustamises. Seoses sellega, et fookus pole enam moraalsel eksimusel, vaid kahju põhjustamisel, kaob interpersonaalse süüteooria kontekstis moraalilise märksõna süü määratlusest ära. See võib tunduda tervitatava arenguna, eriti üldisemate trendide taustal, mis rõhutavad seda, et moraalifilosoofidel on kombeks olnud emotsioone üle moraliseerida

(D'Arms & Jacobson, 2000; Prinz, 2009). Näiteks räägitakse tihti vihast kui reaktsioonist ebaõiglusele, kuigi tegelikult võivad inimesed vihastuda ka lihtsalt ärritava heli peale (Prinz, 2009). Kas ei ole sama eksimust tehtud ka süütunde puhul – traditsiooniliselt räägitakse süüst seoses moraalse eksimusega, kuigi tegelikult võivad inimesed tunda süüd ka lihtsalt kahju põhjustamise tõttu? Väidan siiski, et asjaolu, et kausaalse süü juhtumid ei sisalda moraalse eksimuse komponenti, ei tähenda, et moraal nendest sootuks välja jääks. See saab ilmsiks siis, kui vaatame pelga kahjutundmise juhtumeid, mis ei tõuse süütunde tasemele – midagi, millele interpersonaalse käsitluse kohaselt ruumi ei tohiks jääda. Kui näiteks professor annab tudengile halva hinde, siis võib tal küll tudengist siiralt kahju olla, kuid see ei pruugi veel süütunnet tähendada, kui leitakse, et hinne on õiglane. Seda tüüpi juhtumid näitavad, et süütunne ei ole üksnes kahju põhjustamise küsimus. Inimene peab sündinud kahju ka moraalselt halvana nägema. Kui näiteks ollakse autoavariis kellelegi surmavat kahju põhjustanud, siis ei ole tegu lihtsalt moraalivälise juhtumiga. Kausaalse süü juhtumite puhul seisneb moraalne halbuses selles, et leitakse, et juhtunu on ohvri suhtes siiski ebaaus. Täpsemalt on siinkohal tegu mittevõrdleva õigluse valda jääva väärilisuse küsimusega – ohver ei väärinud nõnda halba saatust. Sellest lähtuvalt vaidlen, et isegi, kui moraalsus ei tule kausaalsetesse juhtumitesse sisse toimija ebamoraalse käitumise kaudu, jääb see siiski lahutamatuks süütunde osaks selle kaudu, et olukord, mida põhjustati, on moraalselt halb.

Neljas peatükk keskendub ellujääja süütundele ning laieneb ka üldisemalt juhtumitele, kus inimesel on teiste suhtes ebaõiglane privileeg. Kui pelgalt kausaalse süü juhtumite puhul oli probleemiks see, et käsitlus polnud piisav – jättis moraaliga süüst välja –, siis ellujääja süü puhul on alternatiivsete käsitluste probleemiks see, et nad toovad sisse liiga palju ebavajalikke elemente. Vaatlen seda viga kahe tähtsama süükäsitluse näol, mis püüavad eksplitsiitselt arvesse võtta ka just ellujääja süüd. Esimene neist on endiselt interpersonaalne süükäsitlus, teiseks David Vellemani (2003) hirmupõhine süükäsitlus. Interpersonaalne lähenemine selgitab kõiki süütunde juhtumeid nende sarnasusega prototüüpsele kahju põhjustamise stsenaariumile. Baumeisteri ja kollegide sõnul on ellujäämisstsenaarium sarnane kahju põhjustamisele selle poolest, et inimene saab teiste arvelt kasu (1994, p. 247). Teiste arvelt kasusaamine on küll igati relevantne Holokausti üle elanute puhul, sest Holokausti kontekstis oli üsna tava-pärane, et gaasikambritesse saatmiseks olid seatud täpsed kvoodid, mille täitmist hoolikalt jälgiti – mis juhul võib tõesti öelda, et kui kellelgi õnnestus pääseda, siis osalt just sellepärast, et keegi teine saadeti gaasikambrisse enne teda. Siiski ei ole teiste arvelt aspekt tarvilik kõikide ellujääja süütunde juhtumite puhul, näiteks olukorras, kus vähist paranenud inimene tunneb ellujääja süüd oma palatikaaslaste ees, kellel ei õnnestunud eluga pääseda. Velleman (2003) soovib paneb ette, et ellujääja süütunnet seob traditsiooniliste süütunde juhtumitega see, et mõlemal juhul kardetakse teiste inimeste õigustatud halvaksapanu, lihtsalt tavaliselt kardetakse teiste vihast halvaksapanu, soovitud kui ellujääja süütunde puhul kardetakse kadedusel põhinevat halvaksapanu. Mõlemal juhul sisaldab halvaksapanu aga ebaõigluse elementi. Väidan, et Vellemanil on küll õigus selle osas, et ellujääja süü

seondub ebaõiglusega, kuid ta toob ebaõigluse süütundesse sisse valel viisil – selle kaudu, et kardetakse teiste ebaõiglusepõhist emotsiooni (halvakspanu). Vaidlen, et süütunne ei saa seisneda teiste halvakspanu kartuses, sest see teeks süütundest egoistliku emotsiooni. Sedavõrd, kuivõrd süütunne on oluline ka andestamise kontekstis, ei taha me endaga halvasti käitunud inimeselt mitte seda, et ta kardaks meie halvakspanu ning muretseks sellepärast, kuidas ennast õigustada, vaid, et ta mõistaks, et käitus meiega ebaõiglaselt. Niisiis peaks fookus olema murel ebaõigluse enda pärast, mitte sellega kaasnevate sanktsioonide pärast. Väidangi, et ellujääja süütunde selgitamiseks ei peaks vaatama kaugemale ebaõiglusest endast ning kaitsen minimalistikku käsitlust ellujääja süüst kui süüst, mille puhul subjekt tajub end konstitutiivse osana teiste suhtes ebaõiglasest asjade seisust.

Viiendas peatükis sõlmin kolmanda ja neljanda peatüki järeldused kokku ühtseks süükäsitluseks, mida nimetan hüvepõhiseks, sest rõhuasetus ei ole mitte ebamoraalsetel tegudel, vaid moraalselt halbadel tagajärgedel. Hüvepõhise süükäsitluse kohaselt seisneb süütunne selles, et subjekt näeb end panustavat moraalselt halba asjade seisust. Kohe selgitan täpsemalt ka panustamise mõistet. Kausaalse süü puhul panustab subjekt halba asjadeseisust kausaalsel viisil, seevastu kui ellujääja süü puhul panustab subjekt halba asjadeseisust konstitutiivsel viisil. Oluline on siinkohal see, et nii kausaalne kui ka konstitutiivne panustamine on sel määral siiski sarnased, et mõlemat saab mõista kontrafaktuaalses tähenduses. Kui subjekt ei oleks teatud kellaajal teatud teel sõitnud, siis ei oleks õnnestust juhtunud. Kui subjekt oleks hukkunud koos teistega, siis olnuks lõpptulemus vähemalt võrdlevas mõttes vähem ebaõiglane olnud. Kõik see rõhutab asjaolu, et subjekt teeb maailma vahe sisse. Antud käsitluse üheks oluliseks vooruseks on see, et seda võib vaadelda teatud laadi keskteena traditsioonilise ning interpersonaalse süükäsitluse vahel. See ühendab interpersonaalsest süükäsitlusest tuleneva idee, et süü on pigem väljapoole kui sissepoole vaatav, traditsioonilise arusaamaga, et tegu on siiski moraalimotsiooniga (olgugi, et moraalilise aspekt ei tule mitte toimija ebamoraalsusest, vaid asjadeseisust moraalsest halbusest). Lõpetuseks rõhutan, et kuigi hüvepõhine süükäsitlus pakub välja igati koherentse viisi, kuidas süütundest saab rääkida, ei peaks seda siiski võtma üldise süükäsitlusena, mis seletab ära kõik mõeldavad süütunde juhtumid. Pigem võrsub sedalaadi süü ühest võimalikust moraalifilosoofilisest vaatest, mille kohaselt moraalilise on keskne headus/halbus, mida saab omistada asjade seisudele, ning tegude õigsus/ väärus on tuletatav sellest, millised asjad seisevad kaasa toovad. Kui sellise käsitluse puhul on fookus hüvel, siis on see igati loomulik, et süütunne väljub deontilistest piiridest ehk läheb ebamoraalsetest tegudest kaugemale. Niisiis näitab käesolev doktoritöö seda, kuidas see, mida süütunne inimeste jaoks võib tähendada, on suuresti mõjutatud nende üldisematest moraalifilosoofilistest vaadetest.

## CURRICULUM VITAE

Name: Heidi Meriste  
Date of birth: 19.01.1990  
Citizenship: Estonia  
Contact: heidy.meriste@ut.ee

### Education

Sept 2015–2025 PhD student of Philosophy, University of Tartu (Estonia)  
April–May 2019 Visiting scholar, University of Geneva (Switzerland)  
May–June 2017 Visiting scholar, University of Konstanz (Germany)  
Sept 2012–June 2015 MA in Philosophy (*cum laude*), University of Tartu (Estonia)  
April 2015 Visiting student, University of Konstanz (Germany)  
Sept 2008–June 2012 BA in Philosophy (*cum laude*), University of Tartu (Estonia)  
Sept 2010–May 2011 Visiting student, Kansai Gaidai (Japan)  
Sept 1996–June 2008 Juuru Gymnasium (Estonia), gold medal  
Sept 2006–July 2007 King Edward’s School Witley (UK)

### Career

2013–... Assistant at Centre for Ethics (University of Tartu)  
2020–2023 Junior researcher at Centre for Ethics (University of Tartu)  
2015–2018 Committee member at the national student research competition (Estonian Research Council)  
2008–... Jury member and lecturer at the Estonian Philosophy Olympiad

### Main research areas

Theory of emotion, moral psychology, normative ethics

### Publications

Meriste, H. 2019. Against Exclusively Retrospective Guilt. In: Cokelet, B. & Maley, C. J. (eds), *The Moral Psychology of Pride*. Rowman & Littlefield, pp. 71–93.

### Talks

*Interpretivist Developments in the Philosophy of Emotion*. Workshop “Interpretivism, mental fictionalism and folk psychology” (November 8, 2024, Tartu).

*Vastutusest ja süütundest Krossi „Kajalugudes“ (Antiphoni ja Bernard Williamsi taustal)*. The Annual Estonian Philosophy Conference (August 30, 2024, Tartu).

- Modern Guilt Revisited: Footnotes to Bernard Williams*. Workshop “Mind with a Historical Twist” (December 8, 2023, Tartu).
- Critique of the dualist account of guilt*, the 10<sup>th</sup> Annual Conference of the European Society for the Study of Emotions (June 15–17, 2023, Tartu).
- In Defense of a Broad Monist Account of Guilt*, the Conference of the International Society for Research on Emotion (July 15–18, 2022, Los Angeles).
- In Defense of Character Guilt*, the 9<sup>th</sup> Annual Conference of the European Society for the Study of Emotion (June 9–11, 2022, Graz).
- Guilt and Damage to the Moral Order*, the 8<sup>th</sup> Annual Conference of the European Society for the Study of Emotions (June 10–12, 2021, Zoom).
- In Defense of a Unified Conception of Guilt*, weekly seminar of the Thumos research group in University of Geneva (October 3, 2019, Geneva).
- Guilt, Anxiety, and Normative Vulnerability*, International Society for Research on Emotion (ISRE) annual conference (July 10–13, 2019, Amsterdam)—co-authored with and presented by Juliette Vazard (University of Geneva).
- Differentiating Guilt and Shame*, in summer school “Empathy, Ethics, and Errors” (June 18–21, 2019, Pühajärve).
- Against Exclusively Retrospective Guilt*, University of Helsinki practical philosophy seminar (March 13, 2019).
- Can you feel guilt about the future?*, Humane Philosophy Project / Ian Ramsey Centre’s 2018–2019 seminar series (January 17, 2019, Oxford). See the video: <https://youtu.be/hkQDc0rBx4A>
- Beyond Retrospective Guilt*, Helsinki University philosophy of science seminar (November 12, 2018).
- Past-, Present-, and Future-Oriented Dimensions of Guilt*, the 4th Conference on Traditions and Perspectives of the Phenomenological Movement in Central and Eastern Europe on “Phenomenology of Emotions” (October 12–13, 2018, Kaunas).
- Contempt as a Response to Superbia*, conference “Emotions: Rationality, Morality and Social Understanding” (September 7–9, 2017, Tartu).
- Critical Response to Macalester Bell’s Conception of Contempt*, conference “Negative Emotions: the Good, the Bad, and the Ugly” (30–31 May, 2017, Geneva).
- Normative Analysis of Research Integrity and Misconduct*, Research Integrity Workshop (September 19, 2016, Bristol), co-authored and presented with prof. Margit Sutrop.
- Lying and telling the truth are not the only two alternatives*. Issues in Bioethics and Philosophy of Medicine. Transformation of Patient Information Practices. Workshop of Nordic-Baltic Network for Philosophy of Medicine and Medical Ethics (June 2013, Riga).

### **Teaching Experience**

- 2017/18 Guilt: Conceptual and Moral Aspects (3 ECTS), University of Tartu  
2017/18 Pride: Conceptual and Moral Aspects (3 ECTS), University of Tartu  
2017/18–... Main Topics in Ethics (6 ECTS), University of Tartu  
2016/17 Shame: Conceptual and Moral Aspects (3 ECTS), University of Tartu  
2014/15 Sports Ethics (3 ECTS), University of Tartu  
2014/15–... Basics of Ethics (3 ECTS), University of Tartu

### **Organisation of conferences and workshops**

- June 15–17, 2023 The 10<sup>th</sup> Annual Conference of the European Society for the Study of Emotions (co-organised with Bruno Möelder, and Triin Paaver)  
June 14, 2023 Workshop “Emotions—more like stars or constellations?” (co-organised with Bruno Mölder, and Uku Tooming)  
Sept 7–9, 2017 Conference *Emotions: Rationality, Morality and Social Understanding* (co-organised with Margit Sutrop, Bruno Möelder, Vivian Bohl, and Triin Paaver)  
May 14–18, 2015 International Philosophy Olympiad (co-organised with Leo Luks, and the University of Tartu Science School)

### **Reviewing**

Articles for journals: *Emotion Review*, *Synthese*, *Journal of Didactics of Philosophy*.

## ELULOOKIRJELDUS

Nimi: Heidy Meriste  
Sünniaeg: 19.01.1990  
Kodakondsus: Eesti  
Kontakt: heidy.meriste@ut.ee

### Hariduskäik

Sept 2015–... doktoriõpingud filosoofias, Tartu Ülikool  
Aprill–mai 2019 külalistudeng Genfi ülikoolis (Šveits)  
Mai–juuni 2017 külalistudeng Konstanzi ülikoolis  
Sept 2012–Juuni 2015 magistrakraad filosoofias (*cum laude*), Tartu Ülikool  
Aprill 2015 külalistudeng Konstanzi ülikoolis (Saksamaa)  
Sept 2008–juuni 2012 bakalaureusekraad filosoofias (*cum laude*), Tartu Ülikool  
Sept 2010–mai 2011 vahetusüliõpilane Kansai Gaidai ülikoolis (Jaapan)  
Sept 1996–juuni 2008 Juuru Gümnaasium, kuldmedal  
Sept 2006–juuli 2007 külalisõpilane King Edward’s School Witley’s (Suurbritannia)

### Teenistuskäik

2013–... Referent Tartu Ülikooli eetikakeskuses  
2020–2023 Nooremteadur Tartu Ülikooli eetikakeskuses  
2015–2018 Komisjoniliige õpilaste riiklikul teadustööde konkursil (Eesti Teadusagentuur)  
2008–... Žüriiliige ja koolitaja riiklikul filosoofiaolümpiaadil

### Peamised uurimisvaldkonnad

Emotsiooniteooria, moraalipsühholoogia, normatiiveetika

### Publikatsioonid

Meriste, H. 2019. Against Exclusively Retrospective Guilt. Kogumikus: Cokelet, B. & Maley, C. J. (toim), *The Moral Psychology of Pride*. Rowman & Littlefield, lk 71–93.

### Ettekanded

*Interpretivist Developments in the Philosophy of Emotion*. Töötuba “Interpretivism, mental fictionalism and folk psychology” (8. november 2024, Tartu).  
*Vastutusest ja süütundest Krossi „Kajalugudes“ (Antiphoni ja Bernard Williamsi taustal)*. The Annual Estonian Philosophy Conference (30. august 2024, Tartu).

- Modern Guilt Revisited: Footnotes to Bernard Williams*. Töötuba “Mind with a Historical Twist” (8. detsember 2023, Tartu).
- Critique of the dualist account of guilt*, the 10<sup>th</sup> Annual Conference of the European Society for the Study of Emotions (June 15–17, 2023, Tartu).
- In Defense of a Broad Monist Account of Guilt*, the Conference of the International Society for Research on Emotion (15.–18. juuli, 2022, Los Angeles).
- In Defense of Character Guilt*, EPSSE (European Society for the Study of Emotions) aastakonverents (9.–11. juuni, 2022, Graz).
- Guilt and Damage to the Moral Order*, EPSSE (European Society for the Study of Emotions) aastakonverents (10.–12. juuni 2021, Zoom).
- Ellujääja süütunne ja selle süüvabad analoogid: piirsituatsioonide retrospektiivsest moraalifenomenoloogiast*, Eesti Kirjandusmuuseumi kirjandusteaduse X suvekool “Kirjanduse (po)eetiline ruum” (21.–22. august 2020, Peedomäe).
- In Defense of a Unified Conception of Guilt*, Genfi ülikooli Thumose uurimisgrupi igapäevane seminar (3. oktoober 2019, Genf).
- Guilt, Anxiety, and Normative Vulnerability*, International Society for Research on Emotion (ISRE) aastakonverents (10.–13. juuli 2019, Amsterdam), kaasautor ja ettekandja: Juliette Vazard (Genfi ülikool).
- Differentiating Guilt and Shame*, suvekool „Empaatia, eetika ja vead“ (18.–21. juuni 2019, Pühajärve).
- Empaatilise süütunde anatoomiast Jaan Krossi loomingus*, Eesti kirjandusmuuseumi IX suvekool „Kirjanduse empaatiline hää“ (14.–15. juuni 2019, Juhan Liivi Museum, Alatskivi).
- Against Exclusively Retrospective Guilt*, Helsingi ülikooli praktilise filosoofia seminar (13. märts 2019).
- Can you feel guilt about the future?*, Humane Philosophy Project / Ian Ramsey keskuse 2018–2019 seminarisari (17. jaanuar 2019, Oxford). Vt ka videot: <https://youtu.be/hkQDc0rBx4A>
- Beyond Retrospective Guilt*, Helsinki ülikooli teadusfilosoofia seminar (12. november 2018).
- Past-, Present-, and Future-Oriented Dimensions of Guilt*, the 4th Conference on Traditions and Perspectives of the Phenomenological Movement in Central and Eastern Europe on “Phenomenology of Emotions” (12.–13. oktoober 2018, Kaunas).
- Süü- ja häbitunde seostest Jaan Krossi „Kajalugudes“*, Eesti uuringute tippskeskuse kevadkonverents „Olgem eurooplased, aga saagem ka eestlasteks” (27.–28. aprill 2018, Rakvere).
- Contempt as a Response to Superbia*, konverents “Emotions: Rationality, Morality and Social Understanding” (7.–9. september 2017, Tartu).

*Critical Response to Macalester Bell's Conception of Contempt*, konverents "Negative Emotions: the Good, the Bad, and the Ugly" (30.–31. mai 2017, Genf).

*Kas alandlikkus ja kahetsus on olemuslikult halvad?*, Eesti filosoofia aastakonverents (11.–13. mai 2017, Kääriku).

*Normative Analysis of Research Integrity and Misconduct*, Research Integrity Workshop (19. september 2016, Bristol), koos prof. Margit Sutropiga.

*Lying and telling the truth are not the only two alternatives*. Issues in Bioethics and Philosophy of Medicine. Transformation of Patient Information Practices. Workshop of Nordic-Baltic Network for Philosophy of Medicine and Medical Ethics (June 2013, Riga).

*Kas eksitamine on vähem halb kui valetamine?* Eesti filosoofia aastakonverents (oktoober 2012, Tallinn).

### **Õpetatud kursusi**

2017/18 Süü: kontseptuaalsed ja moraalsed aspektid (3 EAP), Tartu Ülikool

2017/18 Uhkus: kontseptuaalsed ja moraalsed aspektid (3 EAP), Tartu Ülikool

2017/18–... Eetika peateemad (6 EAP), Tartu Ülikool

2016/17 Häbi: kontseptuaalsed ja moraalsed aspektid (3 EAP), Tartu Ülikool

2014/15 Spordieetika (3 EAP), Tartu Ülikool

2014/15–... Eetika alused (3 EAP), Tartu Ülikool

### **Konverentside ja töötubade korraldamine**

15.–17. juuni 2023 Euroopa Emotsioonide Uurimise Filosoofilise Ühingu 10. Aastakonverents (koos Bruno Mölderi ja Triin Paaveriga)

14. juuni 2023 Euroopa Emotsioonide Uurimise Filosoofilise Ühingu 10. Aastakonverentsi eelne töötuba *Emotions—More Like Stars or Constellations?* (koos Bruno Mölderi ja Uku Toominguga)

7.–9. sept 2017 Konverents *Emotions: Rationality, Morality and Social Understanding* (koos Margit Sutropi, Bruno Mölderi, Vivian Bohli ja Triin Paaveriga).

14.–18. mai 2015 Rahvusvaheline filosoofiaolümpiaad (koos Leo Luksi ja TÜ Teaduskooliga).

### **Artiklite retsenseerimine**

Ajakirjades *Emotion Review*, *Synthese*, *Journal of Didactics of Philosophy*.

## DISSERTATIONES PHILOSOPHICAE UNIVERSITATIS TARTUENSIS

1. **Jüri Eintalu.** The problem of induction: the presuppositions revisited. Tartu, 2001.
2. **Roomet Jakapi.** Berkeley, mysteries, and meaning: a critique of the non-cognitivist interpretation. Tartu, 2002.
3. **Endla Lõhkivi.** The sociology of scientific knowledge: a philosophical perspective. Tartu, 2002.
4. **Kadri Simm.** Benefit-sharing: an inquiry into justification. Tartu, 2005.
5. **Marek Volt.** The epistemic and logical role of definition in the evaluation of art. Tartu, 2007.
6. **Aive Pevkur.** Professional ethics: philosophy and practice. Tartu, 2011.
7. **Toomas Lott.** Plato on Belief (*doxa*) *Theaetetus* 184B–187A. Tartu, 2012, 208 p.
8. **Jaanus Sooväli.** Decision as Heresy. Tartu, 2013, 153 p.
9. **Ave Mets.** Normativity of scientific laws. Tartu, 2013, 217 p.
10. **Vivian Bohl.** How do we understand others? Beyond theories of mind-reading and interactionism. Tartu, 2014, 184 p.
11. **Uku Tooming.** The Communicative Significance of Beliefs and Desires. Tartu, 2014, 208 p.
12. **Andrus Tool.** Objektiivsuse teema Wilhelm Dilthey vaimuteadustefilosoofias. Tartu, 2014, 329 p.
13. **Janar Mihkelsaar.** Giorgio Agamben and Post-Foundational Political Ontology. Tartu, 2015, 169 p.
14. **Jaana Eigi.** The Social Organisation of Science as a Question for Philosophy of Science. Tartu, 2016, 282 p.
15. **Tarmo Tirol.** Subjektiivsus Jean-Paul Sartre'i eksistentsialismis. Tartu, 2017, 253 p.
16. **Eve Kitsik.** Revisionary Ontology: Improving Concepts to Improve Beliefs. Tartu, 2018, 196 p.
17. **Juhan Hellerma.** Mapping Time: Analysis of Contemporary Theories of Historical Temporality. Tartu, 2020, 138 p.
18. **Henrik Sova.** The Structure of a Consistent Global Pragmatism. Tartu, 2020, 87 p.
19. **Riin Kõiv.** The Content and Implications of Nativist Claims. A Philosophical Analysis. Tartu, 2021, 182 p.
20. **Lisa Zorzato.** Towards a Realist Approach to the Challenge of Fictional Models: Augmented Determination. Tartu, 2023, 150 p.
21. **Jay Zameska.** The Ethics of Public Health: Balancing the Interests of Populations and Individuals. Tartu, 2023, 122 p.
22. **Semen Reshenin.** Individual Self-Determination in Friedrich Schiller's Philosophical Writings. Tartu, 2023, 228 p.

23. **Sergei Sazonov.** The Entrepreneurial Theory of Ownership. Tartu, 2024, 191 p.
24. **Juho Lindholm.** Knowledge and Practice. A Criticism of Epistemology. Tartu, 2024, 372 p.