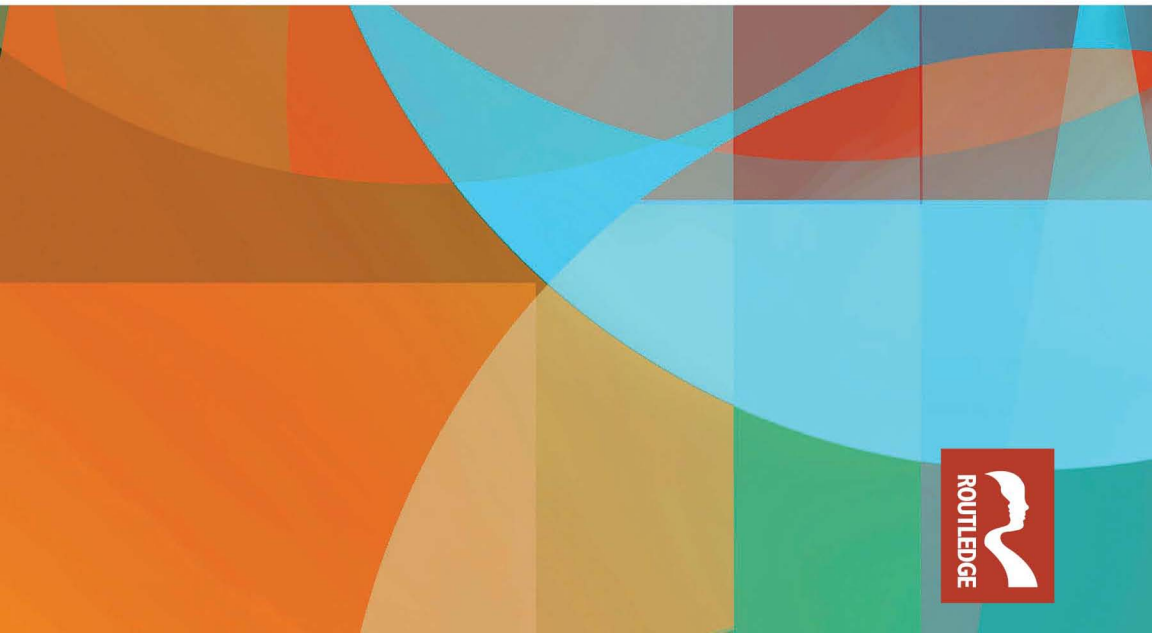
Routledge Studies in Ethics and Moral Theory

UNFAIR EMOTIONS

THEIR MORALITY AND BLAMEWORTHINESS

Jonas Blatter



Unfair Emotions

This book provides a novel philosophical account of the unfairness of certain emotions. It explains how the concept of unfairness can be applied to emotions and how emotions can be the proper objects of second-person moral evaluation.

Emotions are an integral part of our moral practices. While the links between emotions and morality have received much philosophical attention recently, the phenomenon of unfair emotions remains under-explored. This book examines an everyday phenomenon that we often perceive other people's emotions as unfair, in a similar way as if they acted unfairly. It argues that the notion of unfairness combines elements of the unfittingness and of the moral relevance of an emotion. In the first half of the book, the author shows how an unfair emotion can wrong another person. His account holds that an emotion is unfair to its target if its inherent action tendencies constitute a directed moral hazard to the targeted person. In the second half, the author examines to what extent we are responsible for feeling an unfair emotion, and in what way we can – and cannot – be held accountable for it. He argues not only that emotions can be unfair but also that there are limits to when we may hold people accountable for them.

Unfair Emotions will appeal to scholars and graduate students working in ethics, philosophy of emotion, moral psychology, and cognitive psychology.

Jonas Blatter is a philosopher working on the intersection between ethics and philosophy of emotion, with a focus on moral responsibility and interpersonal norms. After his PhD at the University of Bern, he became a Postdoc researcher at Ruhr University Bochum, expanding his research to include interactions with emotive AI.

Routledge Studies in Ethics and Moral Theory

Risk and Responsibility in Context
Edited by Adriana Placani and Stearns Broadhead

Moral Thought Outside Moral Theory
Craig Taylor

Virtuous and Vicious Expressions of Partiality
Edited by Eric J. Silverman

Responsibility Collapses
Why Moral Responsibility is Impossible
Stephen Kershnar

A Phenomenological Analysis of Envy
Michael Robert Kelly

The Self, Civic Virtue, and Public Life
Interdisciplinary Perspectives
Edited by Nancy E. Snow

Moral Agency in Eastern and Western Thought
Perspectives on Crafting Character
Edited by Jonathan Jacobs and Heinz-Dieter Meyer

Unfair Emotions
Their Morality and Blameworthiness
Jonas Blatter

For more information about this series, please visit: www.routledge.com/Routledge-Studies-in-Ethics-and-Moral-Theory/book-series/SE0423

Unfair Emotions

Their Morality and Blameworthiness

Jonas Blatter

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

First published 2025
by Routledge
605 Third Avenue, New York, NY 10158

and by Routledge
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

ISBN: 978-1-032-80674-7 (hbk)
ISBN: 978-1-032-80673-0 (pbk)
ISBN: 978-1-003-49803-2 (ebk)

DOI: 10.4324/9781003498032

Typeset in Sabon LT Pro
by Apex CoVantage, LLC

For my friends and family
I'm not mad if you don't read it, though


Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Contents

Preface *ix*
Acknowledgements *xi*

1 Introduction 1

1.1 What Are Emotions? 2
1.2 The Phenomenon of Unfair Emotions 4
1.3 The Problems and Challenges 10
1.4 The Benefits 12
1.5 Overview 14

2 Criticism of Emotions 17

2.1 Unfittingness 23
2.2 Inconsistency 30
2.3 Prudential Criticism 33
2.4 Moral Criticism 37
2.5 Towards Unfairness Criticism 42

3 Unfair Emotions 45

3.1 Morally Objectionable 48
3.2 Conditionality 57
3.3 Interpersonal Directedness 64
3.4 Directed Wrongs 67

4 Responsibility and Control 74

4.1 The No Control Problem 75
4.2 Disentangling the Control Requirement 77
4.3 Controlling Emotions 80
4.4 Ways of Holding Accountable 87

5 Accountability for Emotions 100

5.1 Holding Others to Expectations 102
5.2 Moral Versus Non-Moral 104
5.3 Co-Valuational Expectations 108

6 Fair Resentment for Unfair Emotions 117

6.1 Distinctions 118
6.2 Resentment as Ill Will 123
6.3 The Aim of Anger 126

7 Conclusion 137

7.1 Outlook 138
7.2 Upshots 141

Bibliography *144*
Index *150*

Preface

Moral responsibility and how we hold others to account have always been a fascinating topic to me. 'What am I to blame for?' seems like a foundational question of moral philosophy that is almost as basic as 'What should I do?' However, the discussion of responsibility is often inextricably linked to the debate about whether we have free will, one of the grand old, ever-burning, and irresolvable debates of philosophy. I did probably share my frustration with this stalemate with people like Peter F. Strawson, whose approach to responsibility in the form of reactive attitudes has deeply influenced my thinking and writing on the topic ever since I read 'Freedom and Resentment'.

It was R. Jay Wallace's take on the reactive attitudes in 'Responsibility and the Moral Sentiments' that inspired my thinking about emotions like anger and resentment being the proper objects of moral evaluation. While Wallace does present an in-depth analysis of the discussion of the fairness of the entire practice of holding people responsible, the analysis of fairness has always left me unsatisfied. If the practice of holding responsible, meaning being disposed towards reactive emotions like anger and resentment, can be unfair, then surely the emotions arising from such a disposition should themselves be considered unfair. But in later writings, Wallace adopted the notion of *fittingness* and *warrant* to characterize the standard by which reactive emotions should be judged. These notions, stemming from the philosophy of emotions literature, lack precisely the sense of moral force that a criticism of *unfairness* so powerfully suggests.

Building on my initial hunch that the locus of unfairness in our practices of moral responsibility should be found in the reactive emotions themselves, I decided, for my PhD thesis, to find a more morally loaded alternative to the notion of *fittingness*. Such a notion should be able to capture the sense of wrongfulness that people, like me, sometimes feel when someone unjustifiably gets angry with them. However, the phenomenon is not limited to reactive emotions but seems more broadly to be possible for any

emotion that is directed at another person and has a negative tinge to it, such as fear or disgust.

The hardest part of this project has turned out to be selling the idea to people and especially to other philosophers. The notion of being morally judged on something that is both very personal and intimate as well as not voluntary, such as an emotion, is at first both frightening and upsetting. It seems itself unfair to be judged on something that is not under my voluntary control, and it feels like an invasion of privacy to have an aspect of one's inner life morally scrutinized. But counterintuitively, this reaction has motivated me even more to pursue this project. I see a real need to talk about this issue for two reasons. First, despite it feeling invasive and mean-spirited, people judge each other based on their emotions all the time. This is an aspect of social life that will not disappear anytime soon. This leads to the second reason, to not talk and write about when and how it can be justified to judge someone's emotion as unfair means simply to suppress and deny that we do it. To resolve the matter by deeming it irrational does not get rid of the phenomenon itself.

This book is the product of my research into the unfairness of emotions. It is not the first to discuss the morality of emotion, but there is still a lack of discussion around this topic, in philosophy, but also in other disciplines, like psychology, and probably also in many people's everyday lives. My hope is that even if people disagree with the specific account I offer or the points I make, that this can be a contribution to a debate that deserves more attention. It touches many other topics. As mentioned, it is an issue I think lies at the heart of the Reactive Attitudes account of moral responsibility. It is certainly also a sensitive issue in questions of love, friendship, and other close relationships. But the fairness of how we feel towards others is also an interesting topic in and of itself. I hope the full discussion in this book can convince some of my readers.

Acknowledgements

I want to thank Andreas Müller and Fabrice Teroni for supporting my doctoral studies and research as my supervisors. I have benefited greatly from your expertise in the fields that informed this book, your constructive feedback, and the many discussions that informed my thinking on issues of morality and emotions.

My fondest thanks also go to my fellow PhD students, Stephanie Elsen, Rodrigo Diaz, Annabel Colas, Delphine Bracher, Yusuf Yuksekdag, Melanie Altanian, Kelly Tuke, Stephanie Deig, and the regular participants of the practical philosophy colloquium, Lukas Nägeli, Marcel Twele, Sabine Hohl, Andreas Cassee, Matthias Rolffs, Anna Goppel, and Markus Stepanians, with whom I led the many invigorating discussions in the halls and spaces of the University of Bern. It is these interactions that make a university a place that inspires thought and creativity, and makes it into a place I want to be.

My most grateful thanks also go to all the people who took the time to discuss my work with me in conferences, workshops, seminars, or in person, including the participants in the Thumos seminar with Julien Deonna and Fabrice Teroni; the participants of the Oberseminar of Sabine Döring; the members and affiliates of the Ethics Centre in Zürich and the BBZ workshops with Peter Schaber; the participants of our 'Ethics of Emotions' workshop, including Macalester Bell, Justin D'Arms, Leonhard Menges, Anne Meylan, Jonathan Mitchell, Sebastian Schmidt, Laura Silva, Christine Tappolet, Mauro Rossi, Vida Yao, Daniel Telech, Rachel Achs, Katherine Rickus, Samuel Reis-Dennis, and Per-Erik Milam; and the many others who have counselled and inspired me along the way, including Constant Bonard, Juliette Vazard, Maude Ouellette-Dubé, Dong An, Stefan Riedener, Paulina Sliwa, Thomas Schmidt, and R. Jay Wallace.

My most heartfelt thanks go to Tanja Rechnitzer, my partner in life and in staggering through the labyrinth that is academia. I am immeasurably thankful that you were there for me through the many challenges that led to this book, and all the other ups and downs in our intertwining lives.

My research and studies have also greatly benefited from the courses, community, and support provided by the Swiss Doctoral School in Affective Sciences at the Swiss Center for Affective Sciences, University of Geneva.

Last but not least, my thanks go to the Swiss National Science Foundation for enabling this project with their financial support for the open-access publication of this book, and for funding my research as part of the project 'Reasons and Emotions' (Grant number: 189011).

1 Introduction

If I get angry with you over nothing, you will probably resent me for it. There seems to be an everyday phenomenon just like this – that we respond to other people's emotions with a type of moral disapproval. We even seem to sometimes perceive them as *unfair*, similarly to unfair actions. For example, when I get angry with you for no good reason, you might get upset yourself and feel seen in an *unfair* light, or when you feel fear of a stranger due to the colour of their skin, you might feel *guilty* for it. We know responses like getting upset more typically from situations where we are either the victim of an offence or when we are wronged by someone, and we know feeling guilty from when we ourselves have done something morally wrong. That similar reactions seem common in response to mere emotions like anger and fear raises the question whether those emotions can themselves be unfair.

From the perspective of moral theory, this might initially seem misguided or even morally objectionable, since we don't usually consider emotions to be the type of thing that is open to moral criticism or even blame. It is actions and overt behaviour that can be unfair and constitute moral wrongs, but surely not emotions. But this initial appearance is misleading, as I will argue in this book. My main thesis is that emotions can indeed sometimes be unfair, and even that it is sometimes appropriate to blame people for such unfair emotions. Moreover, moral theory already has many of the tools and concepts required to understand why and when emotions can be unfair, and even when we are to blame for them, and when they are blameless mistakes. In this book, I will show how we can use those resources to formulate a theoretical account of unfair emotions.

The aim of this introductory chapter is to clarify and motivate the goal of the book and to illustrate the major problems that have to be overcome to achieve it. Broadly, the main aim is to find a conception of what unfair emotions are that can be supported by theoretical positions in ethics and the philosophy of emotions. At the same time, I aim to avoid problems

DOI: 10.4324/9781003498032-1

that are well-established and discussed in the literature on the nature of emotions – such as the moralistic fallacy.[1]

In Section 1.1, I briefly introduce the concept of an emotion and how it is usually understood within current philosophical debates. I also introduce some of the most prominent philosophical approaches to the notion of an emotion, and delineate emotions from other, closely related concepts like moods and character traits.

In Section 1.2, I give a brief introduction to the kind of cases of unfair emotions that I will focus on in this book. In Section 1.2.1, I discuss two possible philosophical approaches to this everyday phenomenon, namely, either to debunk it as irrational or even immoral, or to develop a theoretical account that can accommodate it in our general moral theory. In Section 1.2.2, I give some initial motivation why I take the latter approach and discuss unfair emotions directly instead of only focusing on closely related phenomena, such as the expression or the actual behavioural consequences of emotions.

In Section 1.3, I present the major problems we face when trying to account for unfair emotions in moral theory. The first problem is that emotions, at first glance, don't seem to constitute a moral wrong in and by themselves. The second problem is the apparent lack of control we have over our emotions. In Section 1.4, I present some benefits that an account of unfair emotions might bring both in general for the understanding of our social interactions with respect to emotions and theoretically for certain debates in the philosophical literature – such as debates around reactive attitudes (RA) theory. Finally, in Section 1.5, I give a brief overview of how I will proceed in the book and how I approach solving the presented problems.

1.1 What Are Emotions?

Emotions are an aspect of our minds that are at least as central and important to our lives as intentions or beliefs. But what exactly they are is a point of much debate. While there is disagreement about what elements are constitutive of emotions, most philosophers and psychologists agree that there are a number of features that are closely associated with emotions.

These include, firstly, bodily and hedonic feelings. Bodily feelings include, for example, sensations of an increased heart rate or respiration, sweating, tensing of muscles, often face movements, but also less clearly locatable sensations of elation or heaviness. Hedonic feelings on the other hand mainly refer to whether an emotion feels good, like joy or relief, or bad, like grief or shame. One of the most influential early psychological accounts of emotion, by William James, identifies emotions with just such bodily feelings.[2] However, this feelings account of emotion has since fallen somewhat out of favour. While it is a topic of discussion whether there can

be emotions without feelings, it is less controversial that certain feelings are very commonly involved with emotions.

A second commonly acknowledged feature of emotions is that they involve a focus on an object, event, or state of the world, in the past, present or potential future. This aspect is often used to distinguish emotions from other, albeit closely related, states like moods.[3] An emotion like fear is focused on something like a bear in front of you, or the effects of climate change on human society. In comparison, you could also simply be in a fearful mood, which is not directed towards anything in particular, but a state in which you are generally apprehensive or cautious about nothing in particular. Certainly, moods can and often do spawn emotions, for example, when your generally fearful mood focuses on a specific potential threat, we might say that it manifests in an emotion.

As a third feature, emotions are said to be evaluative mental states. This means that they do not simply focus on an object, but they also present that object in a specific evaluatively charged way. For example, while fear presents something as a threat or danger, anger instead presents its object as an offence or affront. This is a commonly proposed approach to both understand what emotions are and differentiate different types of emotions. Some theories go so far as to entirely identify emotions as mental states that are at their core evaluative judgements about their objects.[4] For example, fear is the judgement that its object is dangerous, or anger is the judgement that its object is offensive. Some versions of this approach take issue with comparing or identifying emotions with judgements, but instead propose to view them as a type of or closely similar to perception.[5] This is, for one, because emotional evaluation seems much more open to being overridden by other beliefs. For example, if you know that a given spider is not dangerous, then you don't take your fear of it as relevant as you would otherwise.

A fourth commonly mentioned feature of emotions is their close association with specific types of behaviour or motivation towards a type of behaviour. Fear, for example, is closely associated with behaviours like fleeing, hiding, or avoiding any potential threats or dangers, while anger is more associated with confrontation and attacking a nuisance heads-on. Just as with the above features, there are also approaches to understanding emotions that identify emotions primarily with this motivational feature.[6] This motivation or tendency towards characteristic types of behaviour does not need to be understood to necessarily manifest in overt actions, but can also be thought of as mental types of behaviour. For example, in grief or sadness, the behaviour might be more in the way a person thinks about their loss, and inability to focus on anything else. The specifics of how this motivational tendency manifests will differ strongly depending on the type of emotion in question.

In addition to these specific features of emotions, there is also a vagueness in what exactly it is when we refer to emotions. For example, there are at least two different ways to understand a sentence like: 'Kai is angry with you'. For one, it could mean that Kai, at this moment, is feeling a muscle tension and elevated heart rate, is focused on you and what you did wrong, and is being motivated to confront it head-on. This is typically called an episodic view of emotion or an emotional episode. For another, we could understand it as meaning that Kai, while not at the moment thinking about you, did not yet forget what you did, and if confronted with the fact again, would experience all of the above. This, in contrast to the episodic view, is a dispositional view of emotion, or an emotional disposition. Both views can either be understood as what emotions really are, while viewing the other as a distinct while related phenomena, or they can be seen as two aspects that together constitute a more complex phenomenon we call emotions.

There are some related, but distinct, types of dispositions that are sometimes referred to in emotional terms. For example, by 'Kai is angry', we could also mean that Kai is generally an angry person, meaning that he tends to become angry very quickly. This is a character trait of Kai that is closely related to emotions, rather than being an emotion in and of itself. Further, we can think of certain attitudes that do involve the disposition towards experiencing certain emotions, which are nonetheless not themselves emotions in the dispositional sense.

The phenomenon I want to focus on in this book are emotions understood as these episodes of increased feeling, that focus on specific objects in an evaluative way, and that inherently involve characteristic motivational tendencies towards types of behaviour. While the dispositional aspects of emotions are certainly important, they can be treated as derivative of emotional episodes. Whatever an emotional disposition disposes to is what is inherent to the emotional episode, hence if we think that the emotional episode itself would be unfair, then a disposition towards it would mean being disposed to something unfair, while conversely, if we find the disposition unfair already, then certainly that is on the basis of what it disposes to.

1.2 The Phenomenon of Unfair Emotions

That emotions can be a proper object of moral evaluation is a widely acknowledged, albeit still controversially discussed, phenomenon in the philosophical literature.[7] However, less attention has been devoted to the types of moral criticism made against certain emotions. In the following, I want to focus on one specific type of criticism, which I call *unfairness*. While this is a type of moral criticism, not just any moral criticism of emotions falls under the category of unfairness. To get a picture of the distinction I have in mind, consider the following examples.

On the one hand, not all morally objectionable instances of an emotion are *unfair*. For example, if your friend is in a deep existential crisis and in the course of that throws some deeply hurtful and offensive insults at you, it would be perfectly fair to be angry with them because of that, but it might also be a bad thing to get angry, given that you want to save your friendship and support your friend. It would still be fair to get angry, if your friend is despite their crisis an overall rational agent and responsible for their behaviour. It seems like you could not be blamed for getting angry at their behaviour, since being in crisis does not necessarily excuse targeted offensive behaviour like that. But, it might not be very advisable, or even morally justified, to feel angry in this situation. This is especially the case if we think that there are special duties of loyalty and care that we have towards friends, like supporting them in a crisis and helping them get through it.

On the other hand, not all unfair emotions are all–things–considered the morally wrong thing to feel. For example, it might be unfair to get angry with the customer service hotline worker because you are frustrated with the service of some larger company. But if the company is sensitive to a large amount of negative, angry customer feedback, it might help to lead to overall positive change. In this case, it is unfair since the worker is clearly not the one personally at fault for the bad service and therefore does not deserve your anger. That specific worker is simply the unlucky representative of the larger company. But the relative discomfort the worker has to endure might be justified if it leads to an overall positive change for a large number of customers.

In both of these examples, the specific overall moral prescriptions can be a matter of disagreement. But what they are meant to show is that we can make a conceptual distinction between an overall moral assessment of an emotion and a more local assessment of what I have been calling its *fairness*. The fairness of an emotion seems to be both a matter of moral evaluation but still separable for the overall moral evaluation of the emotion. It is a *moral* matter, since it makes sense to talk about the service worker not deserving your anger. But it is separable from the *overall* moral evaluation, since we can pose both the question 'Is this emotion overall morally admissible?' and 'Is this emotion fair?', and the answers to both can fall apart – as in the examples above. This strongly suggests that there is a type of cases in which emotions are unfair that is distinct from a general moral evaluation of the emotion. In this book, I want to focus on this type of moral criticism.

1.2.1 Philosophical Approaches

There are at least two philosophical approaches to how to deal with a phenomenon like the one outlined earlier: (i) to try and find a theoretical

account that can be incorporated into our larger moral theory and provide criteria for when the criticism that an emotion is unfair is justified and when it is not or (ii) to discount the phenomenon by either (a) debunking it or (b) arguing that we should reform and stop criticizing emotions as unfair altogether. A debunking strategy (a) would be to show that something else is going on in the above examples; that what we actually do in these situations is judging people for expressions of their emotions, but not the emotions themselves. A reform strategy (b) would be to show that criticizing emotions as unfair is either irrational or even immoral.

While I suspect that the instinct of many philosophers is to pursue approach (ii), I want to pursue approach (i) in this book. I develop a positive account of unfair emotions in the main part of the book. In the remainder of this section, I provide some initial motivation for why we should try to account for the phenomenon and against discounting it. In the following, I provide two lines of argument against a debunking strategy (a). In Section 1.2.2, I argue that there is a good case to be made that in the relevant cases, we should focus on the emotions and not any closely related phenomena, to pinpoint what is morally objectionable in the situation. In Section 1.2.3, I argue that the debunking strategy faces more challenges than we might initially think, especially when it comes to disentangling emotions from their immediate expressions. To reject strategy (b) would mean to show that it is not always irrational or immoral to criticize emotions as unfair. This is achieved by making the positive case for (i), that we can give an account of unfair emotions within a moral theory framework, which is the task of the rest of this book.

1.2.2 Why Focus on Emotions?

The question at hand is whether we should even start to pursue this project, or whether emotions should remain outside the realm of moral criticism. When we think about what we should do, morally speaking, we don't normally think that we should feel a certain way. In the same vein, we don't think that what we can expect of others includes their emotions. 'Don't tell me what to feel!', seems to come more easily than 'Don't feel like that!' When we hear someone tell others how they should feel, many objections may come to mind. For one, what we feel seems our own personal matter. To dictate what people are allowed to feel seems like a violation of this intimate domain. For another, how can anyone expect from us to feel or not to feel on the spot? Our emotions are so often overwhelming and not at all under our control. We cannot be expected to be under a moral duty to feel if we cannot willingly comply with it. Despite such initial intuitions to the contrary, I propose that emotions should be considered a valid subject of moral evaluation and criticism. I claim that the mental state

of experiencing an emotion itself can constitute a moral wrong against another person. To show how such a claim can be defended requires a bit of groundwork. In this introductory chapter, my argument remains at a relatively high level of generality.

In the following, I first want to establish that the emotions of other people are something in which we all have a valid interest. Because mental states like emotions can have a considerable impact on our interpersonal relations, we should take them seriously. They constitute a considerable part of our personal relationships, as well as the basic moral relations of any two people. Not all mental states are equally relevant to others. There are clearly attitudes that have little to no impact on our relationships, and attitudes that have considerable impact. To illustrate this point, we can imagine two similar scenarios that differ in this aspect.

Your Cake: There is one last piece of cake in the fridge and knowing that you would like to eat it, I have told you that it is there. An hour later, the piece of cake is gone, and I assume that you have eaten it.

My Cake: There is one last piece of cake in the fridge and, not having had any yet, I told you that I would really like to eat it later. An hour later, the piece of cake is gone, and I get angry with you, assuming you have eaten it.

In both scenarios, you have in fact not eaten the piece of cake but someone else did, and I was unjustifiably jumping to the conclusion that you did. In *Your Cake*, you may criticize me for not being a very thorough investigator and that I might do better not to jump to unwarranted conclusions too quickly. In *My Cake*, you may do the same but you may also understandably resent the accusation implicit in my reaction.

If the impact of an attitude is what makes it relevant for moral criticism, it raises the question why we should focus on the attitude and not just on the actual impact. If emotions impact a person's thinking and motivation in a way that poses certain hazards to others, it might be best to focus on whether those manifest in any concrete actions. Rather than caring about whether I get mad, maybe you should only be concerned whether I sufficiently suppress my anger or mitigate its effect on my behaviour.

An alternative explanation of what happens in the *Cake* cases is that it is not actually a problem that I get mad at you, but that it would be one only if I express this emotion to you. That is to say, if I did not show my anger or some form of aggressive or confrontational behaviour towards you in any way, you would have nothing to complain about. However, if I did express my anger or treat you differently based on it, you would be in the right to resent my unwarranted behaviour. Certainly, the latter is not very controversial and showing unwarranted confrontational behaviour can be

criticized. But does this mean that overt behaviour and expression are the only relevant aspects of the situation?

To clarify, I acknowledge that *expression* can be understood in a narrow sense as only referring to acts of communication. These can include speech acts or other acts that are clearly recognizable by others as conveying a certain meaning. But *emotion expression* can also be understood in a wider sense. While it includes all the cases of the narrow sense, it also refers to the wider impact of an emotion on a subject's behaviour. This includes facial expressions, approach to or avoidance of the target, acts of social exclusion, or the amount of attention one gives to someone. In the following, I will use the wide sense of *emotion expression* to discuss whether it should be our main focus of moral evaluation, or whether emotions can be morally evaluated directly.

There is a methodological issue to how we interpret the situation. We commonly use fictional scenarios and thought-experiments like the above *Cake* cases to illustrate that there is a moral problem here. However, in the *My Cake* scenario, even if we share the intuition that there is something morally problematic about the scenario, it is not clear what it is. We might have different opinions on which aspect of the scenario elicits our intuition that there is a moral problem. I think that getting mad in this situation is what's unfair, but you might think that it is actually the negligence that leads to me getting mad which is morally problematic, or even that there is nothing *morally* problematic unless I express my anger or let it impact how I treat you.

There are nonetheless good reasons for why we might want to focus our criticism on emotions themselves. Not only can hiding your anger be detrimental to your relationships, if it comes out, it can also be considered a morally questionable form of deception. Rather, assuming that we still interact with each other and I act completely normally, I am pretending that everything is fine. If you suddenly find out that I am in fact angry with you for eating the cake, it would not be out of place for you to be surprised and ask 'why didn't you say anything and act like everything is normal?' But if only expression matters, then such a reaction would not make sense.

One could argue that if it never comes to light that I was secretly mad at you, then you don't have any cause to complain. After all, my anger did not affect your life in any way. However, this makes some questionable assumptions about what morally matters to us – in particular, that something only matters for our moral evaluations if it impacts the actual experience of a person. Imagine that all of your life were like this. Not only am I secretly mad at you, but never let you know it, your family and friends all despise, pity, or resent you but never let you know it. Would you want to lead this kind of life, in blissful ignorance of how people really feel about you?[8] If you are like me, you find this scenario horrible and would prefer

to know – or rather that people wouldn't feel so dreadfully about you. This supports the claim that it matters to us how people actually feel, and not simply what they express.

I don't think these differences in interpretation can be resolved by intuitions alone but need to be argued for on a theoretical level. Here, I merely argue against a basic scepticism about the moral criticizability of emotions. This does not amount to a full theoretical account of how and why emotions can be morally problematic. Rather, I defend the aims of the project to develop an account of unfair emotions and make the case for why we need such an account.

1.2.3 An Uncertain Distinction

So far, I have provided some reasons for why we would want an account of unfairness that focuses on the emotions themselves rather than closely related phenomena, like their expression. In the following, I also formulate some arguments against the debunking approach – the idea that we should focus only on other factors such as expression or the behavioural consequences of the emotion.

My first counterargument against focusing on the expression or consequences is that such a strategy relies on an implicit distinction between the emotion and its impact on the subject, and that this distinction might be a spurious one. This counterargument can be made by either a weaker or a stronger claim. The weaker claim would be that it is not practically possible for a person feeling an emotion to remain fully unaffected by it. Even if I try to hide my anger and do a good job of it, I could hardly ever fully negate all effects that my anger has on my behaviour. Subtle effects might still always creep back into my behaviour. For example, encountering a chance to be generous to you, I am more likely to not do so given that I am still secretly angry at you. It is certainly right that there are better and worse ways of expressing an emotion like anger. Being slightly biased against you is clearly better than falling into a violent rage, attacking you, or scheming in secret against you. But it remains a problem, even if consistently held in check. If this weaker claim is true, then a problematic emotion will always co-occur to problematic forms of behavioural influences. While we can still make the conceptual distinction, in practice, the question whether only the influence or also the emotion is morally problematic becomes far less pressing.

The stronger claim is that an emotion by its very nature consists in part of a kind of behavioural influence. On the one hand, an emotion itself can plausibly be understood as a series of changes to a person's bodily and mental processes. Consequently, most involuntary expressions of emotion, like facial expression and physiological changes, are components of that

change and can therefore be considered part of the emotion. Moreover, the changes to a person's mental processes are inextricably tied to those very physiological changes. Increases or decreases of hormones like dopamine, adrenaline, or serotonin not only are responsible for the bodily changes but also impact our mental processes. To deliberate what to do while *on anger* is already a different kind of behaviour than doing so while emotionally unaffected.

There are different ways to make this point, which depend on different theories of emotion. It can be argued that not only physiological and mental changes are components that make up an emotion but also that changes to overt behaviour are just as much a part of the emotion itself. If an action necessarily involves certain changes to mental states that are affected by an emotion – like motives, intentions, or desires – then experiencing such an emotion means performing different actions. If some mental processes like deliberation can be classified as mental acts, then emotions don't only *affect* but *are* changes in how we act.

If you suppress an oncoming bout of deep sadness and don't let it affect you, do you still experience the same emotion but only reduce its impact on your behaviour? Or do you change the emotion itself, resulting in a different, namely, weaker affective state than you would have experienced otherwise? I don't think it makes sense to claim that what you experience is still the same emotion. Certainly, the phenomenological aspect and bodily feelings are different. Letting the emotion swell up to the point where you are heaving and in tears feels different from keeping it down and focusing on something else. You are less likely to think of the source of your sadness and perceive it as a terrible thing. Hence, you do not let the evaluative component become as salient as it otherwise would. The behavioural impact is lessened, which is probably one of the most common reasons why one would suppress an emotion – aside from its sheer unpleasantness. All these components combined result in a noticeably different affective state which cannot be the same emotion as if it were not suppressed. Given this close connection, criticizing the strong effect your emotion has on your behaviour is the same as criticizing the strength of your emotion, and vice versa.

1.3 The Problems and Challenges

I now want to introduce some of the problems that a philosophical account of unfair emotions faces, which I will address more fully over the course of the following chapters. There are good reasons for why we would want to limit moral criticism only to people's emotional expressions and not allow for emotions to directly be the object of criticism. Limiting moral criticism in this way simply avoids all the potential problems that morally criticizing emotions seems to have.

Many forms of emotion expression can be adequately described as actions and therefore under a person's voluntary control – with the exceptions of involuntary facial expressions and physiological changes. Indignation or resentment can be expressed in speech acts, for example, by accusing or blaming the person one resents. I could express my anger by demanding restitution, or indirectly by not showing you the same goodwill as I normally would. These are all overt, tangible forms of behaviour that express an emotion and for which I can clearly be held accountable. Consequently, there is no need for special treatment of these forms of expression and most conventional moral theories can account for them.

One problem, faced by an account of unfair emotions, is to define what exactly can be morally objectionable about an emotion itself, and not simply about its expression or behavioural consequences. It seems that an emotion can neither harm or otherwise impact another person's well-being, nor can it limit their autonomy or freedom. What kind of interference or relevance can an emotion by itself pose towards anyone but the person feeling it? If we cannot plausibly find any such morally objectionable feature, then emotions cannot by themselves be unfair.

Additionally, unfairness implies not only that there is something morally objectionable about the emotion, but also that it is unfair *to* someone – a *directed* moral wrong. An unfair emotion *wrongs* the person at whom it is directed. But this raises the further difficulty of how to explain this directedness of the moral wrong inherent to an unfair emotion. The simplest way to show that there is such a directedness would be to show that the target of the emotion has a right against such an emotion, which they can claim against the subject feeling the unfair emotion. However, it is implausible that we actually have rights to or against what other people should feel. If it cannot be shown that there is a directedness to the moral wrong of an unfair emotion, then we could not fully account for the initial appearance of the phenomenon that emotions can be unfair *to* their targets. I address this first set of problems in Chapter 3.

A further problem is that it seems we are not responsible for our emotions in the way we would be responsible for our actions or overt behaviour. So even if we can account for the unfairness of an emotion, it initially seems that we should never criticize the subject of the emotion in any stronger way than to simply point out the unfairness. A large part of the apparent phenomenon is, however, that when people feel an unfair emotion towards someone, they are *being unfair* to that person. This implies that they are, at least in some circumstances, *at fault* for the unfairness. Hence, the type of reaction that seems appropriate is not only to criticize the unfair emotion, but also to blame the subject, or hold them accountable in some way. But if we are not in control of our emotions, then how could we ever be accountable? This seeming dilemma has to be overcome if we

want to account for, not only the moral problem with unfair emotions, but also answer the question whether we are ever to blame for them. I address this problem in Chapter 4 and then go on to develop a positive account of accountability for unfair emotions in Chapter 5. Finally, I defend my account against a couple of further challenges in Chapter 6.

1.4 The Benefits

While my project faces the problems and challenges introduced in the last section, if it succeeds, it will also have some substantial benefits. If we denied the possibility of genuinely unfair emotions and limited the scope of what can be morally criticized to actions, we could avoid the mentioned problems. However, I maintain that these problems can be overcome and that there are clear benefits to allowing for moral criticism of emotions: Namely (a), there are advantages for philosophical theory, specifically the reactive attitudes (RA) theory of moral responsibility; and (b), there are desirable practical consequences for how we should see our own and other people's emotional lives. In the following, I elaborate on point (a) for a bit and return to point (b) a bit further down.

On point (a), developing a theoretical account of the phenomenon of unfair emotions is of theoretical interest. Specifically, the question of when it is fair to resent someone lies at the heart of reactive attitudes (RA) theory. Strawson's introduction of RA theory was framed to address the long-standing debate on whether people are morally responsible for their actions whether we lacked free will – due to the world being deterministic or free will being otherwise impossible – or not.[9] He introduced a shift in focus to this debate, away from the question whether people are responsible to the question whether it is appropriate to hold people responsible in the manner we commonly do. According to Strawson, the common way we hold people accountable centrally includes the RA – we get angry with people, or resent them for wronging us, but also feel gratitude for those who help us, and admiration for those who behave virtuously.

A large question at the heart of this turn is how to interpret the notion of *appropriateness* for these reactive attitudes. Some authors argue that the proper standard of appropriateness here is that of *fittingness*.[10] Fittingness is usually understood in one of two ways, either as a standard of correctness or as a standard of epistemic justification. Both interpretations include the idea that types of emotions have certain inherent fittingness conditions that relate them to corresponding evaluative properties. For example, the fittingness of fear depends on the dangerousness of the target. I will elaborate the notion of fittingness in Chapter 2, Section 2.1. In this approach, the question becomes whether it is fitting or epistemically justified to respond with reactive attitudes to people's conduct. This is certainly

an interesting question and worth all the investigations that go into it, but it only addresses a weaker version of the worry about whether we should resent someone for their actions.

The weaker version of the worry is whether it is fitting or epistemically justified to hold people morally accountable by way of the reactive attitudes. Even if the answer to this question is negative, there might still be good reasons for why we should nonetheless continue our practice of doing so. The stronger version of the worry, however, is whether it is morally permissible, or fair, to get angry or resent people for their actions. If the answer to this question is negative, we have a much larger burden to continue our practice. Since whenever we get angry or resent someone despite it being unfair, we would violate a moral requirement. Having an account of when it is unfair to feel an emotion towards someone can shed light on this stronger version of the worry. Namely, it would allow us to ascertain under what conditions it is not only fitting or epistemically justified but also *fair* to get angry or resent someone.

I also want to shortly elaborate on point (b), the benefits an account of unfair emotions would have for our general understanding of emotions in interpersonal relations. While it may not be a philosophically central point, I do think that it is important to understand the practical ramifications of moral theory for our own lives. This is especially true if we adopt moral theories that run counter to folk psychology or everyday moral concepts. It might at first seem like an unreasonable burden to be criticized and take responsibility for what we feel. But if put into context, it can also be a source of empowerment. As philosophers and psychologists[11] alike have argued, it can be empowering for a person to feel a certain ownership over their emotions. It has even been argued that people believing that they are in charge of their own emotions tend to be better at regulating them and display more positive emotions.[12] On the other hand, if you see your emotions almost like external forces that afflict you without your own involvement, you may in effect be less able to regulate them. Nonetheless, feeling ownership can also be disempowering if it leads to feelings of guilt for one's own emotions if they are experienced as wrongful. For example, Lorde points to the feelings of guilt that are instilled in Black women for their own anger in the face of racist and sexist oppression and urges her readers to overcome it.[13] Here, it is the open discussion of emotions as expressions of one's own agency within the social and political context that enables overcoming the potential downsides of taking responsibility for one's emotions.

Responsibility can also be a source of empowerment to be able to criticize other people's emotional states – although this needs to be done with just as much care and caution. There is a thin line between criticizing someone's unfair emotions and invalidating legitimate problems or

grievances. We always need to be mindful of the circumstances in which they occur, but this is just as much the case when criticizing people's behaviour. People are often quick to use emotional states to excuse or justify their behaviour – like the pre-emptive use of violence out of fear. While doing so can in many cases be a legitimate way of enacting one's legitimate cares and interests, it can also be abused. This is especially the case if the underlying emotion, which is put forward to justify a certain behaviour, is itself morally problematic – like racist or bigoted fear. The ability to morally criticize unfair emotions takes away such routes of justification and allows the people affected by emotionally driven bad behaviour to insist on being treated better. Of course, one can also insist on being treated more fairly without morally criticizing someone's emotions that underlie their behaviour. However, it is not clear that this would be a less burdensome demand. It would amount to saying 'yes, of course you're allowed to feel like this, but bottle it up, okay!' – which might under some circumstances be a legitimate response – while criticizing an unfair emotion directly addresses the underlying problem.

1.5 Overview

In the following chapters, I proceed in broadly two steps. In the first half of the book, I establish what aspect or feature of an unfair emotion is morally wrong in the specific sense of unfairness. In the second half of the book, I examine questions about our responsibility for unfair emotions – and whether the subject feeling the unfair emotion is ever to blame for it.

In Chapter 2, I compare the type of criticism involved in judging an emotion to be unfair to other, more common types of criticism of emotion – such as *unfittingness*, *inconsistency*, *imprudence*, or different types of moral criticism. I show how unfairness differs from these types but also what features they have in common.

In Chapter 3, I then formulate my own theoretical account of unfairness, specifically tailored to the discussion of unfair emotions. I argue that A's emotion E is *non-comparatively* unfair to B if E is directed at B, E poses a moral hazard to B through its inherent action tendencies directed towards B, and E lacks intrinsic moral justification – meaning that E's fit-making conditions don't apply or don't offer *pro tanto* moral justification for the E's action tendencies towards B.

From this discussion, it seems that by feeling an unfair emotion towards B, A *wrongs* B, and that B has a special type of complaint against A – that, at least sometimes, B can blame A for E. In Chapter 4, I discuss whether we can ever be to blame for an unfair emotion if we are not responsible for them because we lack direct voluntary control over them. I propose that the answer depends on the conception of blame we use. We can

disambiguate the control requirement of different types of holding someone morally accountable – sanctions and punishment require a different type of control as purely a negative moral appraisal. I argue further that a plausible account of blame, given by the RA theory of moral responsibility, only requires *rational control* and that we can exert that kind of control over our emotions. I thereby show that at least under one plausible account of blame, we are sometimes to blame for our emotions.

In the remainder of the book, I examine how an RA account of blame applied to unfair emotions would look like. In Chapter 5, I specify the conditions under which we are justified to hold someone accountable for an unfair emotion, namely, when that unfair emotion reflects underlying attitudes of disregard for the target. I argue that there are general moral expectations of what kinds of attitudes people should have towards their fellow human beings, which I call attitudes of *basic human regard*. An unfair emotion can reflect a violation of these expectations, and that is when reactive emotions like anger and indignation are appropriate responses to it.

In Chapter 6, I discuss some possible objections to this picture of moral accountability for unfair emotions. First, I argue that the reactive attitudes don't themselves constitute a violation of the expectation of basic human regard, which would pose the threat that my account justifies an infinite circle of blame. Second, I discuss the aims and action tendencies of anger, and whether those render anger an unfitting response to mere emotions. I argue that while anger can lead to aggressive or violent behaviour, blind payback is not the aim of anger. Rather, it motivates confrontational behaviour which can very much be justified in the face of someone showing a lack of basic human regard.

In Chapter 7, I draw some conclusions from what I have shown in the book. I argue that its project was successful in providing a viable and fruitful account of unfair emotions and our moral accountability for them. I end on what I see as its main positive upshot.

Notes

1 See D'Arms and Jacobson, "Sentiment and Value."
2 James, "What Is an Emotion?"
3 See, e.g. Deonna and Teroni, *The Emotions*, 4.
4 Nussbaum, "Emotions as Judgments of Value and Importance"; Solomon, "On Emotions as Judgments."
5 Prinz, *Gut Reactions*; Tappolet, *Emotions, Values, and Agency*; Döring, "Seeing What to Do."
6 Frijda, *The Emotions*; Scarantino, "The Motivational Theory of Emotions"; Deonna and Teroni, "Emotions as Attitudes."
7 See Nussbaum, *Anger and Forgiveness*; Nussbaum, "Transitional Anger"; Bell, *Hard Feelings*; D'Arms and Jacobson, "The Moralistic Fallacy."

8 Nozick, "On the Randian Argument"; Nozick, *Anarchy, State, and Utopia*; and Nagel, "Death" formulate versions of the argument that we are not merely interested in the appearance of people having positive relationships to us, but also their genuine sentiments.
9 Strawson, "Freedom and Resentment," 1993.
10 E.g. Wallace, "Trust, Anger, Resentment, Forgiveness."
11 See, e.g. Gross, "The Emerging Field of Emotion Regulation"; Barrett, *How Emotions Are Made*.
12 Tamir et al., "Implicit Theories of Emotion."
13 Lorde, "The Uses of Anger."

2 Criticism of Emotions

There are many ways in which we can assess and criticize the appropriateness of people's emotions. Almost anyone will have, at some point in their life, said or heard sentences like 'Don't get angry with the children, they don't know any better yet', 'Don't feel intimidated by my parents, they're really nice people', or 'You should be grateful when someone helps you without expecting anything in return'. These are all sentences criticizing an emotional response – or lack thereof – towards others. They all seem to say something similar about the emotion they target, namely, that it is in some way or another inappropriate or that the person feeling the emotion is getting it wrong in some way. They differ, however, in what way the emotion gets things wrong. The first sentence seems to reserve the possibility that anger would be appropriate, were it not for the children still being very young, while the second probably claims that nobody should be intimidated by that person's parents, no matter their age. And the third sentence already seems to gesture towards a norm which governs the adequacy of gratitude. In the philosophical literature, we find many of these different types of evaluations and criticism of emotions, all of which claim that an emotion falls short of some type of norm or adequacy condition. What kind of norms or conditions these are varies depending on the type of criticism. Since, in this book, I am interested in a specific type of criticism of emotion – *unfairness* – I want to first provide some context and distinguish it from other possible types of criticism that could be made about an emotion. This way, we can more easily set aside other notions of appropriateness of emotions and better focus on the question of their fairness.

The type of criticism that I want to examine is a second-personal moral criticism, which I will refer to as *unfairness*. To illustrate the type of unfair emotion I want to analyse, imagine the following example:

> You are waiting in a queue for a ticket counter. At the front, a nervous looking teenager, probably travelling alone for the first time, is trying to clear up an issue with their ticket, which is taking a while. Since this is

DOI: 10.4324/9781003498032-2

> the only open counter, the person behind the teenager is getting noticeably angry at things taking so long. You stand in the queue directly behind the angry queuer, who clearly takes the delay as a personal insult, even giving you a can–you–believe–this–crap look. You can easily imagine what it must be like for the teenager, having to urgently solve this ticket issue, all the while feeling the angry stare in your neck.

What I want to claim about the angry emotion in this scenario is that the queuer's anger is not only irrational or misinformed in some way but also *unfair* to the teenager. We could identify several flaws in the queuer's attitude, such as pointing out that the teenager does not mean any personal insult, or that getting angry does not speed up the process. These types of criticism hint at the false information the emotion is based on or the lack of its prudential value, but neither of them speaks to whether the anger is fair to the teenager. I take this evaluation of the anger as unfair to be a specific type of second-personal *moral* criticism of which we can have a pre-theoretical understanding and that can be distinguished from other types of criticism. To see whether this pre-theoretical distinction holds up under theoretical examination, in this chapter, I examine the differences between unfairness and other forms of criticism, which have been more extensively theorized, and then develop a theoretical account of unfairness in the next chapter.

To clarify these differences, I provide an overview of the most commonly made types of criticism against certain emotions and highlight how they differ from the criticism of unfairness. I focus on the types of criticism which are most commonly found in the philosophical literature, such as unfittingness, inconsistency, prudential criticism, and common types of moral criticism. I show that these types of criticism differ in what they target about an emotion, what standards they apply, their normative implications and the social context in which they are made, such as who can make them. These differences show that all the common types of criticism don't amount to the criticism of unfairness – the second-personal moral type of criticism illustrated in the example above.

To get a more systematic hold on the differences between these types of criticism and unfairness, throughout this chapter, I want to focus on the following four questions applied to each type of criticism discussed: (1) What standard or norm do criticized emotions fall short of? (2) What is the target of the criticism? (3) What is the normative force that the criticism exerts on the subject? (4) Can anyone make the criticism, or is there special standing involved? I propose these questions because they outline four aspects in which the more commonly discussed types of criticism differ from each other and from the criticism of unfairness. While some types of criticism differ from unfairness in all of these aspects, some also share

similarities to unfairness, which will even become useful in the next chapter, where I formulate my own account of unfairness. Nonetheless, my aim for this chapter is to show that unfairness differs from all other types of criticism in at least one of these aspects. Before I go into the other types of criticism, I first want to clarify what I mean by asking these four questions, to give a better sense of what aspect of a type of criticism they are supposed to highlight.

1) What standard or norm do criticized emotions fall short of?

There are many distinct types of standards that we commonly use to evaluate mental states. For example, we typically evaluate our perceptions by how accurately they represent the actual situation before us. This would constitute an epistemic standard of evaluation. In contrast, we usually don't evaluate our actions by such an epistemic standard, but rather by pragmatic or moral standards. That is to say, we either assess how well an action accomplishes a given goal or how well it is justified in light of moral consideration.

Depending on the standard we use to evaluate an emotion, our criticism can have starkly different implications. For example, if I tell you that you don't have to be afraid of my dog because it is very well-behaved, I am simply telling you that your fear is misrepresenting my dog as dangerous, falling short of an epistemic standard of accurate representation. Contrast this to a situation high up in the mountains, where, while crossing over a narrow and slippery path next to a steep drop, I tell you not to be afraid, because fear makes you unsure and leads to mistakes. In this situation, I am not telling you that there is nothing dangerous. Rather, I'm telling you that your fear falls short of a pragmatic standard, it does not misrepresent anything, but it gets in your way of reaching your goal. Like this, different types of criticism can invoke different and sometimes conflicting standards.

What standard does a criticism of unfairness invoke: a standard of correspondence, a principle of rationality, a value not sufficiently realized, or a moral prescription? The unfairness of an emotion could be explained by merely the inaccuracy of how it portrays the world to the subject. If you get angry with me for something innocuous, you are misconstruing the situation in a way similar to being afraid of the good dog. However, it seems like the notion of fairness invokes something more, like a moral standard of evaluating the emotion.[1] Talk of unfairness invokes the notion that someone did not get what they were owed or that they deserved. In the above example, the nervous teenager at the ticket counter did nothing to deserve the ire of the angry queuer. This type of evaluation is also not limited to the so-called *moral* emotions – like anger, indignation, contempt, or guilt. Even fear seems unfair if it is directed at a target who does not deserve to be feared. This becomes apparent in cases of racist or bigoted fear.

2) What is the target of the criticism?

Even when two utterances of criticism sound similar, they sometimes can have very different targets. For example, when you drop and shatter my favourite coffee cup, I could say that your behaviour was very unfortunate or that it was very cruel. In the first case, my assessment mainly focuses on the shattering of the cup itself. I evaluate the event of its shattering or the circumstances that lead to it by some value standard and conclude that the event falls short of being a good outcome. In the latter case, however, I do not only evaluate the shattering as bad, but I imply a malevolent intention or cruel character behind it. Thereby, my criticism also targets either your intentions as morally objectionable or draws your character into question, implying that you are a cruel person.

Criticism of emotions can also differ in what it ultimately targets. For example, if I tell you that your anger is misplaced because you did not know that a traffic accident was the cause for my late arrival, I am primarily criticizing your emotional response as uncalled-for. You could not have known, and you made no mistake in thinking I made you wait for me once again. Contrast this to a situation where I don't help you in moving your furniture because I need to study for an important test, and you get angry. If I then criticize your anger as selfish, I am not only targeting your emotional reaction as inappropriate, I am also implying that there is something wrong with your underlying attitude. In effect, my criticism not only is about your anger but also targets your selfishness.

Then what is typically the target of a criticism of unfairness? Is it simply a statement about the emotion, or does it target the subject as well? Does it imply a flaw in the subject's reasoning, other attitudes, or character? At first, it might look like calling an emotion unfair is a criticism of nothing but the emotion itself and does not also target the subject's character or attitudes. However, we do not usually call things that merely happen, without an agent's intention behind it, unfair. At most, when we do so, we mean to imply that society or the universe has been treating us unfairly. Most types of criticism involve the subjects of emotions to a greater or lesser degree in their assessments of the emotion's appropriateness. Even calling an emotion simply irrational can be said to imply that the subject has made a mistake. At a minimum, the criticism of unfairness targets some form of problematic authorship by the subject – similar to the authorship we have over a terrible idea. In the above example, it is not only that the queuer's anger is unfair, the queuer is also *being unfair* by getting angry at the teenager. It is, in some way, the angry queuer's fault.

3) What is the normative force that the criticism exerts on the subject?

Not all types of criticism call for anything to change. For example, when I tell you that seeing the straw bend at the water's surface is not an accurate representation of the world but merely an optical illusion, I am neither telling

you nor expecting you to correct your visual perception in any way. However, when I criticize your exploitative treatment of your employees as morally wrong, I am very likely implying that you should change it and that you owe them some form of apology or compensation. Hence, negative assessments of actions or mental states can range from something that has no normative upshot, in the case of optical illusions, to forceful accusations that imply either clear expectations of betterment or the threat of sanction or punishment.

Criticism of an emotion as inappropriate can have a similar range of normative implications, depending on what we mean by 'inappropriate', but also on the assumptions we make about the nature of emotions. For example, if I find your enjoyment of what I think is a boring film unintelligible, I am not necessarily implying that you should stop enjoying it. In contrast, when I tell you that your outburst of panic high up in the mountain could have got us all killed, I am clearly implying that you should get a grip on yourself. It is for both your own sake and that of your comrades that you should not let the same kind of emotional episode bubble up again. Hence, you have both rational reason and expectations of others on you to try and regulate your emotions.

What about the criticism that an emotion is unfair to someone? I think it is clear that calling the queuer's anger unfair entails a certain demand to stop being angry. The judgement of unfairness implies that there is something morally amiss or skewed that needs to be corrected, if possible. Calling out someone for being unfair is also a common way to express blame and thereby hold the offending party to account for the unfairness. In sports, when we call the other team's move unfair we typically seek some form of penalty; when a competitor has an unfair advantage, people often demand sanction or fine. In cases like the above, we would rather expect something like an apology or expression of regret. Hence, to invoke unfairness can come with a considerable amount of social pressure and interpersonal expectations. Which makes it even more important to discuss when it might be appropriate to call an emotion unfair.

4) Can anyone make the criticism, or is there special standing involved?

This question does not concern whether other people are capable of determining the fault in an emotion or able to formulate a coherent criticism. Rather, the question is whether making the criticism is only legitimately open to certain people. For example, if you have taken a drink from my refrigerator without asking and drank it, it is up to me to say whether you just stole from me or not. If a third person were to object to your behaviour, I could cancel out any such criticism by making it clear that I do not object. In the above example, it seems that the teenager is in a special position to make such a second-personal complaint that is not easily open to anyone. Even if you as a third person would make the same criticism, you would simply be making it on the teenager's *behalf*.

We can see this in that if the queuer would apologize to the teenager for getting angry, and they accept the apology, then much of the force of the criticism of unfairness would vanish. It might even become inappropriate for you, a bystander, to continue making such a criticism with the same force. It would be similarly strange if you expected the queuer to apologize to you before seizing to criticize. At most, you could continue to judge that the anger was unfair, but the issue is settled now. Or you could judge that you would not so easily accept the apology and would continue making the unfair criticism on behalf of your hypothetical self. In any case, there seems to be someone on whose behalf the criticism is made. The test, to see whether a type of criticism involves this kind of standing, is to ask whether there is an affected person who could in some ways cancel out the force or urgency of the criticism in question.

Other factors that can undercut someone's criticism are hypocrisy or complicity. Does hypocrisy or complicity undermine the criticism of unfairness? In the above case, if the teenager also gets angry every time they have to wait in line when someone is taking too long, that might undermine the criticism of unfairness as well. You could still object that both of them should not get angry in these situations because it is unfair to those other people. But between the angry queuer and the angry teenager, they have much less for which to criticize each other. These are somewhat open questions that I have to address later on. However, the type of criticism I am after is at least open to formulating them in a meaningful way, which is not the case with other types of criticism.

To summarize, the criticism of unfairness seems to have the following features: (1) it entails a moral standard which the targeted emotion falls short of; (2) it targets not only the emotion itself but somehow also involves the subject in its criticism; (3) it typically implies an expectation that something is to be done to remedy the situation; and (4) it is open to questions whether standing or hypocrisy undermines that someone can legitimately make the criticism. In the following, I compare this characterization to other common forms of criticism of emotions, to show how unfairness amounts to a distinct type of criticism.

For the purposes of giving an overview, I distinguish between three broad categories of criticism: Section 2.1, *unfittingness* criticism, which I understand all types of criticism that claim an emotion somehow mismatches the evaluative circumstances; Section 2.2, *inconsistency* criticism, which is concerned with emotions conflicting with beliefs or goals of the subject; and Section 2.4, *moral* criticism, which are all types of criticism that include considerations of an emotions moral worth or reflection on the subject's virtue. In Section 2.5, I draw the conclusion that none of these types of criticism amount to a criticism of *unfairness*, as introduced earlier, and give a brief outlook on what features are missing. In Chapter 3, I then

introduce and develop a type of criticism that is missing from the overall picture, the type of criticism that captures the unfairness of emotions illustrated earlier.

2.1 Unfittingness

The type of criticism most discussed[2] in the philosophical literature on emotions is what I broadly characterize as *unfittingness*. The basic idea uniting the various forms of this criticism is that emotions sometimes do not correctly relate or correspond, in one way or another, to the evaluative properties of their object. For example, fear accurately corresponds to something being dangerous, sadness to something being a loss, and amusement to something being funny. Different emotion theories interpret how emotions need to correspond to these values in different terms, but what they all have in common is that there are instances when emotions incorrectly relate to the object's evaluative properties. For example, when you are afraid of a harmless spider, your fear does not match the actual threat posed by the spider, which renders the emotion open to the criticism of unfittingness. This presents an external criterion for judging the appropriateness of an emotion, based on a correspondence relation between the emotional reaction and an associated evaluative property of the object, the emotion's formal object.

2.1.1 Incorrectness

One way in which the unfittingness criticism of emotions can be spelled out is as *incorrectness* by analogy to false beliefs. Judgement theories of emotion[3] draw this analogy between emotions and judgements by highlighting the evaluative character of emotions. These theories claim that emotions either contain or are themselves a type of evaluative judgement about their objects. In effect, judgement theories liken an incorrect emotion, one that misrepresents its object's evaluative properties, to a false belief or erroneous judgement. This suggests that the subject might be making an epistemic mistake in forming the emotion, similar to drawing a conclusion from insufficient evidence. Alternatively, it might be the case that the person afraid of a puppy is merely misinformed about the capabilities of puppies or otherwise has reason to believe falsehoods that would explain their fear. In that case, their fear would be misplaced and still wrong because they formed it on the basis of unreliable information. In this case, incorrectness criticism would not amount to an epistemic mistake, but merely to the subject being wrong about the situation.

A second interpretation of how emotions can be incorrect is by analogy to false perceptions or illusions. Perceptual theories of emotion[4] view

emotions similar to perceptions, and just like perceptions, they can be more or less accurate representations of the world around us. Perceptual theories frame an incorrect emotion analogous to a false or misleading perception, like seeing a straight stick standing in water as bent where it crosses the water's surface, or as seeing a white chair as red when in a room lit only with red light. Other than misrepresenting colours and shapes, however, an incorrect emotion misrepresents the evaluative properties of its object. Hence, when you are afraid of a harmless puppy, it might seem to you that the puppy is dangerous or that you see it as being dangerous. The criticism that your fear is incorrect then amounts to saying that it misrepresents the puppy as having an evaluative property, being dangerous, even though it does not.

Judgement and perceptual theories have clear ways of establishing an emotion's correctness conditions. They generally hold that emotions have evaluative content, either cognitive content akin to judgements or perceptual content akin to perceptions. Thereby, we can easily establish an emotion's fittingness conditions, namely, when their content correctly represents the object's evaluative properties. However, other theories of emotion like attitudinal theories often also postulate conditions by which unfitting emotions can be criticized as in some way mismatching the evaluative properties of their objects. This third account of the unfittingness of emotions sees it as falling short of a standard that is inherent to the type of emotion in question, as I elaborate in the following.

2.1.2 Epistemic Unfittingness

Deonna and Teroni's *attitudinal theory* of emotion does not ascribe content to the emotions themselves but to an accompanying perception, belief or other type of mental state, like imaginings.[5] This mental state provides its content as a cognitive basis for the emotion and in this way the emotion is about said content. However, the content in question is not necessarily evaluative in itself, rather, it just provides a descriptive representation of the object. In effect, an emotion can be misplaced if the descriptive content of the cognitive basis misrepresents the world, but there is no evaluative misrepresentation involved – as it is with perceptual or judgement theories.

How then can we understand unfitting emotions to misrepresent evaluative properties? Theories like the attitudinal theory nonetheless ascribe *epistemic fittingness*[6] conditions to emotions, which bears some similarities to the correctness conditions of judgement and perceptual theory. While they do not identify emotions with evaluative judgements or evaluative perceptions, they still associate emotion types with specific evaluative judgements. For example, fear does not have the evaluative content that some object is dangerous. However, it is fitting under the same set of circumstances under

which the associated evaluative judgement, that the object is dangerous, would be correct. So while the charge of being epistemically unfitting is not strictly speaking one of incorrectness, it is in many ways parallel.

For one, an emotion can be deemed unfitting in the same circumstances according to attitudinal theory as it could be said to be incorrect according to judgement theories. For another, the fault involved is a similar one, namely, that the emotional response does not match the evaluative properties of the object. In the case of judgement theories, the matching works just like the content of a belief or judgement matching the facts. For perceptual theories, it is akin to how the content of a perception matches the perceived object. In the case of attitudinal theories, the matching is much more independent of the standards of other mental states. An attitudinal theory then portrays fittingness as a relation between emotion types and evaluative properties, governed by the type of attitude the respective emotions are instead of their content.[7] This type of relation can only indirectly be linked to the representational correctness of associated evaluative judgements.

2.1.3 Fittingness Versus Justification

In addition to the fittingness as an analogue to the truth of a belief or the accuracy of perception, we can also distinguish the question whether it is epistemically *justified* – again analogously to justified beliefs.[8] If we are internalists about justification, then an emotion is justified when it is based on the information available to the subject that would make the emotion fitting if it were correct. If we are externalists, then the emotion is only justified if in addition the information it is based on itself is correct.

However, I am not interested here in justified emotions, but in types of criticism of emotions. Hence, the relevant phenomenon would be *un*-justified emotions, meaning instances of an emotion which do not meet the conditions above. The justification of an emotion can fail in at least two ways. Again, if we are internalists about justification it fails if the properties the subject perceives in the object don't and would not make the emotion fitting, and if we are externalists, it can also fail if the subject is mistaken about the object having those properties. Echeverri provides some useful examples of how such a justification can go wrong:

> Suppose that Youna has heard from a reliable witness that her father is in better health. In this case, Youna is permitted to be elated. Crucially, some might want to hold that Youna is still justified in being elated even if the witness happens to be wrong. After all, one might reasonably hold that Youna is permitted to trust testimony in the absence of defeaters. Suppose now that Carlos is afraid of a small spider in the bathroom.

> Yet, his therapist has told him that most spiders in New York City are inoffensive. Remembering this piece of advice, Carlos tells himself: "I should not be afraid of that spider". Yet, the mantra does not work and he keeps shaking. Unbeknownst to Carlos, however, the spider is venomous. In this case, Carlos' fear of the small spider is correct but unjustified.[9]

It seems that in both cases, what goes wrong is either up to a failure of information antecedent to the formation of the emotion – in the case of Youna – or an incoherence of the epistemic agent – in the case of Carlos – which I will cover in Section 2.2. In Youna's case, this might not even amount to the criticism that the emotion itself is unjustified. The criticism is not actually about the emotion at all, but about the belief on which it is based. Here again we have two options. Either we think that Youna's emotion is justified despite her incorrect beliefs – in which case we can at most criticize her for her antecedent beliefs – or that it is only justified if it is based on true beliefs – in which case, the criticism boils down to an incorrectness criticism. In either case, the unjustified status is not actually an additional criticism of the emotion itself.

However, I do not think that the epistemic justification of an emotion is relevant for its fairness. Whether or not an emotion is unfair depends on its fittingness, not whether it appears warranted from the subject's point of view. Hence, the relevant failure is not responding to the perceived situation in a wrong way, but responding to a false perception of the situation. However, questions about the justification of an emotion are certainly relevant for the blameworthiness of the subject, as I argue in Chapter 3, Section 3.2.2. In the following, I therefore focus on unfittingness and not on failures of epistemic justification.

Whichever way one spells it out, unfittingness criticism makes the claim that an emotion does not stand in the right relation with its associated evaluative property, its formal object. For perceptual or judgement theories, this is the case if the object's evaluative properties are not correctly represented by the emotion or the emotion's formal object is not present at all. Attitudinal theories capture more or less the same cases but don't identify the relevant relation between emotion and object as one of representation. In any case, unfittingness as a type of criticism does not fare particularly well with respect to the four dimensions that I have described at the beginning of this chapter, as I discuss in the following.

(1) What standard or norm do criticized emotions fall short of?

The criticism that an emotion is unfitting mainly invokes an epistemic correctness standard. It is epistemic in that what is at stake here is the correspondence to what the object of the emotion is really like. It is not necessarily a standard of good epistemic procedures, but rather one that

is concerned with the quality of the outcome. While emotion theories disagree on the exact nature of the standard involved, it seems a common ground that it is in some way inherent to the emotions. Under what circumstances an emotion is fitting is given either by the emotion's evaluative content or by the emotion type. In any case, moral or prudential reasons for or against an emotion are not part of the epistemic assessment. But in the same way, it does not follow that all fitting emotions are morally good. Being mistaken about a fact or falsely perceiving the situation do not, by themselves, amount to a morally problematic state of affairs. There could, however, be forms of moral criticism that build on notions such as misrepresenting or emotionally mis-responding to the evaluative situation. But we would first need to formulate an additional principle that, for example, *morally* forbids misrepresenting people's moral standing or *morally* forbids certain action tendencies in response to unfitting conditions. Nonetheless, such a principle is not built in to the unfittingness criticism and, in any case, requires further elaboration.

(2) What is the target of the criticism?

The main target of criticizing an emotion as unfitting is the emotion itself, rather than, for example, the rational agent as a whole – as would be the case with incoherence or worries about justification. However, this again differs between the theories of emotion. The incorrectness criticism made by perceptual theorists is mostly directed at the emotions and involves the weakest charge against the subject, if any. It merely states that subjects might be prone to a form of misrepresentation akin to an illusion, a phenomenon that is neither caused by the subject nor able to be dispelled by them. In effect, emotions are something that happens to us, and less something that is strictly speaking our fault.

Something similar can be said of judgement theories. However, there might be more agency involved in making incorrect judgements than being subject to an optical illusion. For example, Nussbaum argues that a subject prone to responding with anger to personal slights betrays a problematic care for personal status,[10] which might implicate a critique of character in her epistemic criticism. However, this is no longer merely the criticism that the emotion is incorrect but includes a further, moral type of criticism. I discuss this type of moral criticism below in Section 2.4. Unfittingness criticism by itself is primarily focused on the emotion and its relation to its object, and less with the subject's abilities or character. While there might be related issues with how the emotion came about, for example, by the subject being inattentive or disposed to make snap judgements, those would amount to additional, different types of criticism.

(3) What is the normative force that the criticism exerts on the subject?

It is unclear whether the criticism of unfittingness comes with any normative force attached. Depending on the exact interpretation, it is akin to

calling something an optical illusion, or it might be similar to a false belief. In the former case, there seems little that could be done about it, most optical illusions are unavoidable, and hence, no normative upshot follows. If we understand an unfitting emotion more like a false belief, there might be some pressure to adapt once we are presented with sufficient evidence. However, it is an open question whether we always have strong reasons to have true beliefs or whether it is simply contingent on how strongly we value truth. It seems relatively clear in cases like fear or anger that you would want to get your evaluative picture of the world right. If you fail to fear a dangerous bear you might get hurt, and if you fear harmless things you're needlessly having a bad time. The same is true for anger, get angry too quickly, and you disgruntle everyone around you, or don't get angry when you should and fail to stand up for yourself or others.

However, there are at least two problems with this approach. First, it is not clear that you can find these types of negative incentives for all kinds of emotions. After all, what would be so bad about being fascinated by a boring pile of mud, or marvel at things everybody agrees to be ugly? There seem to be many instances where positive emotions that erroneously present boring or unremarkable things as pleasant or fascinating can still make your life better or at least make many moments more enjoyable. This leads us to the second problem. Prudential reasons like the ones brought up above, as well as moral reasons, are commonly said to be the wrong kind of reasons to judge an emotion's fittingness.[11] Even if moral or prudential considerations about the object are part of what makes an emotion fitting, this does not make unfittingness a prudential or moral type of criticism. As a comparison, an optical illusion or a false belief can be comforting or pleasing, but they can nevertheless be criticized as incorrect. This all raises the question: 'why be right if being wrong is so much more pleasant?', which I cannot fully discuss here – but which I think goes far beyond the limits of purely epistemic criticism and already involves ethical and prudential considerations, which I return to in the following two sections.

In any case, on the face of it, we don't seem to owe it to anyone specific to change an unfitting emotion. While, if we consider the example above, it seems that at least one reason why the angry queuer should give up their anger is specifically for the teenager's sake. Saying that the anger is unfitting might be right, but that alone does not seem to sufficiently explain why there should be any normative upshot, especially not one that is for the teenager's sake. On the dimension of normative force, unfittingness cannot explain the characteristic force of criticizing an emotion as unfair.

(4) Can anyone make the criticism, or is there special standing involved?

There is nothing intrinsically about the misrepresentation in an illusion or a false belief that is owed to another person. Unfittingness as a criticism is, taken by itself, rather impersonal. Anybody who has access to the

relevant information could criticize an unfitting emotion, independent of whether they are affected by it or have any standing to do so. It is not the case that there is any person who has the standing to cancel out the force of an unfittingness criticism. One might argue that there are people better placed to assess the fittingness of any given emotion, or experts on matters of evaluative properties. However, an expert giving their opinion is not a form of standing, since it does not reliably cancel the criticism of unfittingness. An expert might give you a better epistemic basis to make any such criticism by pointing out features of an object that gives it an evaluative property. But the criticism that an emotion is unfitting is based on just such features, and not the statements of the expert themselves.

It could be argued that it becomes a moral issue when the emotion in question misrepresents the moral status of another person. For example, when you feel contempt for a perfectly nice and overall morally good person. Or it might be a moral issue when an emotion motivates aggression against someone who has done nothing offensive. There is most likely a close connection between what makes these emotions unfitting and their unfairness, but it is not exhausted or explained by the unfittingness alone. We need an additional argument to make the case that we owe it to others not to misrepresent them in this way, or why the resulting action tendencies are morally objectionable. I intend to make such an argument in this book, but it is not already given by the type of unfittingness criticism discussed here.

Other factors, like hypocrisy, that might affect someone's ability to legitimately criticize an emotion's unfairness also don't seem relevant for criticizing an emotion as unfitting. For example, it would be absurd to ask whether being subject to an optical illusion yourself undermines the assessment that someone else's optical illusion is an incorrect representation. So if calling the queuer's anger unfair would simply be a judgement of incorrectness, it should not be affected by the teenager's own tendency to get mad at people taking too long. In the case of a judgement account of emotion, this becomes even more absurd. Hypocritical unfittingness criticism under judgement theory would entail that you criticize another's emotional judgement as incorrect while you yourself would hold the same judgement. This would be akin to telling someone that their belief that the world is flat is wrong, while believing that the world is flat yourself. But this does not amount to hypocrisy, but simply deception or misrepresentation of your own beliefs. Again, to establish such a claim would require an additional moral principle that states when such misrepresentation is morally objectionable.

The best case for unfittingness-based hypocrisy can be made within an attitudinal theory of emotion. If emotions are characterized by action tendencies, then to criticize someone else's emotional reaction while

experiencing the same action tendencies seems close to hypocrisy. However, the criticism involved would still not be as morally problematic as with a moral type of criticism. For example, someone who unfittingly gets afraid all the time could still tell another person that it is unfitting to be afraid all the time. The constantly afraid person might even be in the best place to make such a criticism because they know what it is like to constantly be unfittingly afraid and therefore have learned to recognize it in others. This might be similar to an experienced philosopher telling a student not to be misled by red-herrings or equivocations. The more experienced person might be best to make that criticism because they have themselves fallen into that same trap many times and have learned to recognize it. For these cases to become hypocritical, we need to reinterpret the criticism as involving some type of moral condemnation or assume that the criticizer is making some underlying implications. But unfittingness criticism by itself does not necessarily imply any moral condemnation. Therefore, criticizing an emotion as unfitting due to its action tendencies is also not hypocritical. In cases where there seems to be some condemnation or hypocrisy involved, there is always an additional moral principle involved.

Given the discrepancies between a criticism of an emotion's unfairness and one about its unfittingness in all four dimensions, it is reasonable to conclude that these are distinct types of criticism. This does not exclude the possibility that unfittingness is part of the explanation of unfairness. For example, an emotion being unfitting could be a necessary condition for it to also be unfair. Alternatively, unfittingness paired with a moral principle might be jointly sufficient, even if not necessary, for an emotion to be unfair. I will elaborate on this question in Section 3.2. For now, I turn to another type of criticism of emotions, inconsistency.

2.2 Inconsistency

The second broad category of criticism that can be made against emotions is that they can be in some way inconsistent, either inconsistent with each other or inconsistent with other mental states. The more common version of the inconsistency criticism of emotions is that the emotion in question conflicts with some other mental state of the subject. For example, it could be a *recalcitrant*[12] emotion whereby you feel fear at a spider but at the same time believe it to be harmless. Alternatively, there could be a conflict between an emotional response and your goals. For example, where you feel afraid of flying but need to get to a far-away conference. In both cases, the conflict is not necessarily about the correctness or external validity of the emotion, but merely about the internal conflict between the emotion and another attitude.

An emotion is called *recalcitrant* if it can be said to involve an evaluative representation of the world that is inconsistent with the subject's other evaluative judgements or beliefs.[13] The distinction between this type of criticism and epistemic criticism is that a recalcitrant emotion violates an internal standard of rationality, while epistemic criticism violates an external standard of accurate representation. The internal standard only involves that different mental states should not conflict with each other. Since both judgement and perceptual theories take emotions to have some form of representational content, they can conflict with other mental states that have such content.

(1) What standard or norm do criticized emotions fall short of?

For judgement theory, recalcitrance involves a clear form of rational inconsistency because the judgement involved in the emotion can directly conflict with an independent judgement to the contrary. Take the example of Sam,[14] whose father is a career criminal:

> Sam feels ashamed of his father and believes this to be an appropriate reaction to such a shameful family secret. However, while in college, Sam comes to change his mind and no longer believes that his father's ill reputation reflects badly on him. Therefore, it is not something Sam should be ashamed of. However, Sam still feels ashamed about having such a father, and these emotions conflict with his newfound belief that it is not shameful.

According to judgement theories, Sam's shame includes a form of judgement that having such a father is shameful. At the same time, he now has formed the belief that it is not shameful. Therefore, his emotions directly contradict his explicit beliefs. This is an inconsistency in Sam's mental states that can be criticized as a form of irrationality. Note that this criticism is different from incorrectness, since we don't need to take a stance on whether having a criminal father is actually shameful or not. The inconsistency consists in the conflict between the emotion and the belief whether it is shameful or not. This kind of inconsistency captures an unease about recalcitrant emotions that many authors on the topic seem to share. However, Brady has argued that judgement theory presents this inconsistency as much too strong, and that the conflict between emotions and beliefs is not as severe as having conflicting beliefs.[15]

Perceptual theories present a weaker form of inconsistency between recalcitrant emotions and conflicting judgements. In their view, a recalcitrant emotion is similar to an optical or auditory illusion, like seeing a straight stick bending where it protrudes out of a body of water. When we know that this is an illusion, any reason to believe that the stick is bent is defeated and no real conflict remains. Analogously, having a criminal

father can still seem shameful to Sam, even though he does not believe it to be shameful. In this version, there is no rational criticism involved in the inconsistency. Since having an optical illusion while knowing that it is an illusion is not irrational and recalcitrant emotions are relevantly analogous to such illusions, they are not irrational either.

Brady criticizes perceptual theories' version of the incoherence involved in recalcitrant emotions as too weak.[16] He argues that the irrationality involved is stronger than what perceptual theories allow for, but not as strong as conflicting judgements. He develops his own version, what he calls a neo-judgementalist account of recalcitrant emotions.[17] In his view, emotions are not themselves evaluative judgements but exert a rational pressure to make such judgements. His criticism of recalcitrant emotions is therefore that they rationally pressure us to believe something we should not believe, given our other judgements.

I will not elaborate on the role of recalcitrant emotions in choosing the best theory of emotions here. My aim here is to present the distinct kind of criticism involved in recalcitrance, and how it differs from other forms of mental inconsistency. Just as with incorrectness criticism, recalcitrance assumes that emotions have cognitive content in the form of an evaluative representation of the world. But other than incorrectness criticism, it does not require us to take a stance on whether the emotion is correct or whether the conflicting beliefs are correct.[18] In effect, this type of criticism does not prescribe which way a subject should go, whether they should revise their judgement or overcome the emotion. It is merely the observation that the subject entertains incoherent mental states and that there is some amount of cognitive dissonance that might exert pressure to resolve the conflict.

(2) What is the target of the criticism?

While unfittingness criticism primarily targets the emotion, it is less clear that inconsistency criticism does. In the case of recalcitrant emotions, it might be tempting to identify the emotions as the target of criticism. However, it is not clear in every case that it is the emotion that is creating the problem. In some cases, we should rather resolve the inconsistency by revising our judgement. For example, take a situation where you receive sexual attention from someone that makes you feel uncomfortable. But you judge, due to what your cultural surroundings keep telling you, that it is a positive form of attention that you should be flattered by.[19] This is clearly a case of a recalcitrant emotion of discomfort or discontent, but most people would probably agree that the culprit is the judgement, a type of false consciousness, rather than the emotion. The primary target of the criticism, therefore, seems to be the constellation of emotion and attitude together and not the emotion itself.

In the case of unfair emotions, however, the fault lies clearly on the side of the emotion. There is no judgement that the angry queuer could revise to

make the felt anger less unfair or not unfair at all. That said, a recalcitrant unfair emotion still seems to make a difference in how we would judge the subject. In the example, if the queuer did not judge the teenager taking so long to be worth getting angry over, but nonetheless felt angry, due to an anger problem, we would probably not judge the queuer as harshly. However, this is more a question of accountability of the subject, and less about the unfairness of the emotion itself. I will return to the question of how recalcitrance influences accountability in Chapter 5.

(3) What is the normative force that the criticism exerts on the subject?

Just like unfittingness criticism, criticizing an emotion as inconsistent does not have any moral or interpersonal force to it. Whether someone is being consistent in their mental states is a standard independent of what impact it has on other people. I can criticize you for being inconsistent in your mental attitudes and that might influence you to try to resolve the inconsistency, but you do not in any way owe it to me to do so. The perceptual version of this criticism is even less forceful, since there is no rational pressure to not be subject to an optical illusion. Analogously, if we experience a recalcitrant emotion, according to perceptual theory, we can just acknowledge its incoherence with our other beliefs and go on to faultlessly ignore it.

Under any interpretation of the rational pressure recalcitrant emotions exert, it does not amount to the demand for apology or expression of regret. Being subject to incoherent mental states does not involve a second person, and to argue that we owe coherence to anyone but ourselves would require arguing for an additional moral standard. I don't pursue this option here.

(4) Can anyone make the criticism, or is there special standing involved?

Just like unfittingness, the criticism that an emotion is incoherent with a judgement does not involve other people in such a way that would give them special standing to make the criticism. Anyone who caught on to you experiencing a recalcitrant emotion could point it out equally, and there is nobody who could cancel out other people making the criticism. Also, hypocrisy does not affect anyone's standing to make the criticism.

Given the discrepancies between the criticism of incoherence and that of unfairness, this type of criticism cannot explain the characteristic wrong of unfair emotions. However, it brings up interesting questions, like how the evaluation of an unfair emotion changes if it is recalcitrant or if the subject clearly disavows it.

2.3 Prudential Criticism

Prudential criticism mainly consists in the claim that certain emotion episodes pose an obstacle to or stand in conflict with achieving our goals or with living a good life. An example of the former criticism that emotions

can interfere with our goals is the case where a bout of fear can actually interfere with your ability to deal with a dangerous situation. In such a case, the criticism of fear is not that it is unfitting or incoherent with your belief, you are fully aware that the situation is dangerous, but that it is a hindrance to your immediate goal of re-establishing your own safety.

This type of criticism assumes that emotions have characteristic motivational effects on their subject that can make it harder to do certain things or behave in a certain way. For example, anger motivates more aggressive or confrontational behaviour. In the case of the angry queuer, this might actually lead to the teenager getting caught up in a verbal disagreement and overall take even longer, which runs exactly counter to what the angry queuer would like to happen.

As an example for the latter criticism, that emotions can conflict with leading a good life, Bittner makes the claim that we should never feel regret.[20] This is because regret is an uncomfortable or even painful feeling and on the other side has no benefits that could not be achieved by rational deliberation. Therefore, assuming that there is no point in suffering needlessly, it would be rational to try and rid ourselves of feelings of guilt and become more purely rational with respect to moral guilt.

This criticism of guilt does not require the assumption that emotions come with inherent action tendencies, but only that they have a hedonic quality, for example, that they feel bad. Of course, a similar criticism could be made on the basis of action tendencies. For example, anger could be criticized on the basis that it makes it harder in general to clear-headedly deliberate, and therefore you will often hamper your own attempts to achieve your goals. This has an overall worse effect on your life. And nothing positive comes out of anger that could not be achieved through reasoning. Therefore, you should try to rid yourself of anger completely.

(1) What standard or norm do criticized emotions fall short of?

While the criticism against unfittingness or recalcitrant emotions is based on considerations of theoretical rationality, prudential criticism is based on considerations from practical rationality. It basically states that a certain emotion is not a good means, or is even an obstacle, to achieve a specific end or overall lead a better life. In this, prudential criticism is closer to a moral assessment than the previous types of criticism.

Prudential criticism still focuses primarily on the worth of an emotion for the subject alone. An imprudent emotion is first and foremost an obstacle for the subject's own goals or well-being. To call an emotion unfair, however, necessarily involves the negative involvement of other people in some way. We could widen the scope of prudential criticism to include emotions that can become obstacles to other people's goals and well-being. But that would quickly lead to the question whether we are not actually assessing the emotion's interpersonal moral merit and is no longer merely prudential.

There is a further difference between prudential and unfairness criticism. While an imprudent emotion is simply bad due to its psychological effects, and is so whether it is fitting or not, an unfair emotion only seems problematic if it is also false or misguided. While getting angry at something genuinely offensive can still be imprudent, for example, when showing signs of anger would put you in harm's way, it cannot be unfair. It could still be unfair if you get angry due to a feature of the object that is not actually offensive, while not getting angry about another feature of the object that would actually be offensive. But while this makes a difference for the emotion's fairness, it makes no difference for its prudence. If you find anger useful because you can channel it into cleaning your house, then it matters little what made you angry, nor whether it is fitting – either way it is prudentially useful.

(2) What is the target of the criticism?

Prudential criticism can be directed at a range of different targets. At a minimum, it merely states that this emotional episode would not be conducive to a certain goal. But prudential criticism can also be seen as targeting the subject's tendency to feel such emotional episodes and the overall adverse effects on their live. An overall angry person, for example, might find that their tendency to get angry often gets in the way of what they want or has an overall detrimental effect on their health and happiness. In such a case, the criticism targets the underlying tendency or psychological trait that disposes them to get angry so often.

If we assume that people have a certain influence on their emotional states, prudential criticism can also be directed at the subject of the emotion. This would amount to criticizing someone for getting in their own way, or sabotaging themselves, by fostering certain emotions or maybe simply by letting themselves get emotional at inconvenient times, even if they could stop or suppress the emotion. This kind of emotion regulation would address a prudential criticism since it is not concerned with the emotion's fittingness, but only with the effects it might have on the subject's motivation and psychological state. Put more bluntly, even when you get afraid in actually dangerous circumstances, where fear would be fitting, it might be prudential to suppress your fear.

This focus on the subject also seems to apply to criticism of cases such as the queuer's unfair anger. People typically have a sufficient range of options to regulate their emotional responses, from suppressing them to shifting their attention to something else. These might be enough to make this type of criticism targeted at the subject as an agent, and not merely the inconvenient emotion itself. The angry queuer could simply swallow their anger or seek some sort of distraction from the boring wait in the queue.

(3) What is the normative force that the criticism exerts on the subject?

Pragmatic criticism of emotions mainly appeals to the subject's self-interest. If you want to achieve your goals, this emotion can get in your

way. Hence, if you want to succeed, try to not have emotional episodes of that type. There is no moral force behind this type of criticism, unless it is supplemented with some moral principle that postulates, for example, a duty to self-improvement or to strive to be as rational as possible. But any such demand would be an addition to purely pragmatic criticism. In effect, there is also no-one in particular who could make this criticism more than anyone else or insist on the subject's compliance.

A stronger version of prudential criticism raises the question of why we should ever want to experience certain negative emotions, even when they are correct. Aside from positive emotions that simply feel good, experiencing emotions only seems worthwhile if they have instrumental value. In any case, prudential criticism does not go beyond the subject's own interests. There is no necessary relation to interpersonal sanction or accountability, which would be characteristic of unfairness.

(4) Can anyone make the criticism, or is there special standing involved?

As with prior types of criticism, there is nobody in particular who has special standing to make prudential criticism of an emotion, other than maybe the subject of the emotion themselves. In the specific instance where an emotion poses an obstacle to achieving a goal, the subject could theoretically also decide to give up their goal for the sake of the emotion. For example, imagine you need to get to a meeting soon, but on your way there get struck by the beauty of the evening-sun vista. Your emotion of awe could be criticized as an obstacle to making it to your meeting on time, since it urges you to stop and marvel at the beauty of nature. However, if you decide that the view is worth being late for, you can cancel out this very specific criticism. Barring any further arguments for why you would really want to be on time, this might be an instance where the subject has special standing to make or cancel out a specific criticism.

However, the scope of special standing described earlier seems a rather narrow one. It also does not apply to the type of prudential criticism that certain emotions are a detriment to the good life. It mostly applies in situations where you can easily decide to give up on your prior goals. Also, in the case of unfair emotions, if anyone, it is the target of the emotion and not the subject who has special standing to decide to cancel the criticism in this way.

From the discussion of prudential criticism, it becomes clear that to adequately capture the unfairness of an emotion, we need to look to more clearly moral types of assessment. While prudential criticism captures the practical relevance of emotions, it does so solely focused on the goals and well-being of the subject. Unfairness, in contrast, concerns at least the interests or well-being of another person. I therefore turn next to moral criticism of emotions.

2.4 Moral Criticism

The final category of criticism I want to discuss is moral criticism of emotions. This can include discussions of the harm emotions can do to the subject and others, but also how certain emotions reflect on the subject's moral attitudes or character. We can broadly distinguish two types of moral criticism of emotions: wholesale criticism of specific types of emotions and criticism of specific instances of an emotion.

The first type of moral criticism of emotions, wholesale criticism, is mainly concerned with the negative impact emotions have on human behaviour. A prominent example of this type of criticism in recent years is Martha Nussbaum's criticism of anger. She criticizes anger as an emotion that inherently involves a desire for revenge or retribution.[21] In this characterization, she builds on an Aristotelian view of anger as a response that is primarily concerned with injuries to oneself – or those close and dear – and with injuries to one's relative social status.

Nussbaum argues that anger suffers from a twofold problem.[22] On the one hand, if it is a response to an injury, it presents a form of wishful thinking to seek to remedy the suffered pain through the infliction of more pain to others. On the other hand, if anger is merely a response to a relative down-ranking, it does in fact hit the mark and can be an effective means to re-establish your status relative to the perpetrator. However, in these cases, Nussbaum criticizes the underlying attitude of valuing relative status so highly as misguided and vicious.

Nussbaum's first objection can be reformulated as claiming that anger is not really a fitting response to past injuries. While the painful negative evaluation in anger correctly represents such an injury, the desire for retribution internal to anger is misplaced. This is because getting payback by inflicting a new injury to the offender does neither undo nor alleviate the victim's own injury. The injury inflicted on the other is therefore just more unjustified pain and suffering in the world. Focusing on the second objection, Nussbaum acknowledges that anger is a more fitting response to a denigration of relative social status, but objects that caring for such a status is morally misguided.

Both objections are of a moral nature. The argument that revenge does not undo a past wrong entails the practical criticism that anger seeks an end that is not possible. Since undoing the past wrong is impossible, causing further pain is in no way morally justified, and the desire to do so becomes the desire to act immorally. The second argument, that care for relative status is morally misguided, is a clear moral criticism. Not only does it not further the moral goal of realizing a good life for yourself and others, it actively detracts from it. Her criticism paints all cases of anger as morally deficient, not just specific ones. Even in the case of systematic discrimination, where

she agrees that caring for being socially down-ranked is reasonable, she denies that anger, that is, the desire for retributive down-ranking of the perpetrator, is the morally appropriate response.

A crucial feature of Nussbaum's account is that morally problematic emotions betray an underlying morally problematic attitude or morally dubious care. More precisely, these emotions are only intelligible if we interpret them as rational responses to something that the subject values or cares for.[23] Anger can be warranted from the perspective of the subject when they perceive a slight or down-ranking as an injury. They do so only if they strongly care about their social status, a care which Nussbaum finds misguided.

However, this type of moral criticism does not capture the second-personal nature that I am after. Criticizing someone for overly caring about their social status amounts to criticism of their moral character, not for imposing a morally inadmissible burden on the target of their anger. At most, it involves the criticism that the subject has a vicious character trait that might frequently lead to such burdens on others, but that is not the same as criticizing such an instance for being unfair. The disposition towards second-personal wrongs is not itself a second-personal wrong. Rather, it seems that we first need a clear account for why instances of unfair emotions are morally bad to argue why a disposition towards such instances is a bad thing.

It is also not clear whether it follows that a habitually angry person should be held accountable for their anger. Certainly, we can say that they should not care for their relative status and that the onus is on them to better themselves, but it does not follow that they owe any such remedy to anyone else. Only when their general disposition is instantiated in a specific case of anger that unfairly targets another person do they seem to become accountable to that target, as I will argue in this book. But merely the criticism that someone has such a vicious character trait does not establish any accountability to others.

Other than the wholesale type of moral criticism, discussed earlier, there is a type of moral criticism that discerns individual instances of an emotion and assesses them on their own merits. Typically, moral criticism of emotional tokens is based on a notion of fittingness, as discussed in Section 2.1, but adds a moral principle on top of it. For example, Bell formulates several conditions under which contempt would be morally apt.[24] This marks a defence of contempt against the broad, whole-sale type of criticism, similar to the previously discussed criticism of anger by Nussbaum. But not all instances of contempt are morally valuable, only those where contempt is fitting. Contempt is fitting when it reacts to someone who strongly displays the vice of *superbia*, an attitude of feeling morally superior towards others. By starting with this condition, she builds the

moral evaluation of the emotion on top of a non-moral standard of fittingness. The moral value of fitting contempt can then be defended along the following lines:

First, apt contempt has epistemic value. In this case, the epistemic benefit is not on the side of the subject, but to the target of contempt. Contempt can work as a powerful social signal that you have fallen short of some value or virtue that you hold or that is expected of you. People turning away from you or disengaging can better achieve this than angry engagement could, where you are more likely to try and defend your own point of view against criticism.

Second, apt contempt has motivational value. The motivational character of contempt is to disengage and distance yourself from the target, which can work as a form of protection from destructive or abusive characters. Also, being the target of contempt, and the social shunning that comes with it, can feel terrible and serve as a strong motivation to second-guess yourself. This is morally valuable when it leads to insights about one's own shortcomings and as a potential start for betterment.

Third, apt contempt is intrinsically valuable. There is a line of thought that having attitudes such as valuing the good and *dis-valuing* the bad is in itself morally valuable.[25] In the same vein, contempt presents a way of dis-valuing the target's vices in a specific and matching way. All these sources of moral value are undermined if contempt is unfitting. Showing someone contempt when they did not in fact display the relevant vices becomes epistemically misleading and dangerous. The target would likely get a distorted picture of themselves, thinking they have been subject to a vice they actually don't have. Similarly, unfitting contempt would have bad motivational effects, disrupting potentially valuable relationships and getting people to unnecessarily second-guess themselves. And of course, contempt loses any intrinsic value as a form of dis-valuing something morally bad if it is unfitting.

It is at this point that Bell introduces additional moral principles that establish a moral evaluation on top of a pure fittingness consideration of contempt. In this case, the principle that having a negative attitude, like contempt, towards a similarly negative thing, like something contemptible, is in itself morally valuable. A more general formulation of such a principle might be that it is intrinsically morally valuable to only experience fitting moral emotions – meaning, emotions that have a moral evaluation as their fittingness condition. However, it is not clear what kind of moral criticism such a principle entails. It seems like a principle of moral excellence – that a morally excellent person should only have fitting moral emotions. However, this would not explain any *second-personal* moral criticism against unfitting moral emotions. Nor would it explain the unfairness of any unfitting *non-moral* emotions – like racist fears.

Also, not all of Bell's conditions that render contempt morally inappropriate are about the fittingness of the emotion itself. For one, hypocritical contempt undermines its own value. In such a case, contempt can still be based on a fitting and morally defensible evaluation of its target's vices. But a person who themselves has the vice of superbia cannot defend feeling contempt for someone else showing it.[26] There are other considerations, such as contempt for someone who has developed their vice of superbia under oppressive conditions. Bell views criticism in such a case as unfair, but ultimately leaves it open to further debate under what circumstances the vices of people coming from oppressive backgrounds should be excused.[27] Does this type of moral criticism amount to unfairness? I now turn to comparing the above-discussed types of moral criticism to unfairness, along the four dimensions introduced at the beginning of this chapter.

(1) What standard or norm do criticized emotions fall short of?

Both types of moral criticism discussed in this section are concerned with the moral merits or problems of emotions and go beyond the consideration of fittingness or consistency with other attitudes. In that respect, it clearly matches the notion of unfairness discussed at the beginning of this chapter.

However, wholesale moral criticism cannot differentiate what is wrong with the anger of the queuer from any other instance of anger. By showing the inherent moral problems of anger, we cannot point to any special problem that marks anger as unfair, as opposed to instances when it is not unfair. A wholesale criticism paints the queuer's anger as objectionable in the same way as a protester's anger over political injustice, or the anger of the victim of a crime. A wholesale approach therefore fails to capture the fine-grainedness of the moral evaluation we are looking for.

An instance-based approach seems much more promising in this regard. It employs the idea that the moral value of an emotion is tied to its fittingness conditions, that is to say, to the emotion's formal object. When we generalize this idea to other emotions than contempt, it follows that fear might be morally problematic when it misrepresents someone as dangerous and motivates corresponding behaviour, even amusement could become a problem when it falsely portrays something that happens to a person as funny or comical, rather than, say, a serious accident.

What this type of criticism cannot account for is that what seems to be distinctly unfair about an emotion like the queuer's anger is something directed at the teenager. Linking the moral evaluation of an emotion to its fittingness conditions does not exclude cases where nobody in particular is unfairly targeted. Feeling happy over the destruction of a world cultural heritage site might be a morally problematic emotion, in part because it is unfitting, but it is not necessarily unfair to anyone in specific.

(2) What is the target of the criticism?

Moral criticism also addresses the right kind of target. Moral criticism of an emotion does not typically end at an assessment of the emotion itself. Rather, it additionally connects the defects of the emotion to faults in the subject's character or other dispositions that lead to the morally problematic emotion in the first place.

One issue that remains is that focusing on the subject's character as what is at fault might not be fine-grained enough to account for what we object to in unfair emotions. In examples like the angry queuer, it is not obvious that the queuer is only really unfairly angry if that anger expresses a vicious character trait. It might still be unfair to the teenager if the queuer is otherwise a nice and patient person, and only gets angry due to some adverse situational circumstances. Even if it is not expressing a vice, the teenager does not deserve to be the target of the queuer's anger.

(3) What is the normative force that the criticism exerts on the subject?

On the topic of normative upshot, most moral criticism focuses on the vicious or misguided moral character of the subject or the moral merits of the emotion, rather than the normative consequences for the subject. Neither do these types of criticism usually draw the analogy to blameworthy actions. It is an additional question to what degree the subject can or should be held accountable for their problematic emotions, or whether it is rather a call to morally better themselves. In the latter case, the subject might mostly owe it to themselves to become a better person, or indirectly to all the people around them to become a better member of society.

In contrast, in the case of the angry queuer, whatever else follows from the evaluation that the anger is morally deficient, what should follow is that the queuer owes the teenager something like an apology or expression of regret about getting angry. To put it differently, the normative upshot of something being unfair should involve the party towards whom it is unfair. While this is not incompatible with the discussed types of moral criticism, it does not clearly follow from them either.

(4) Can anyone make the criticism, or is there special standing involved?

A common feature of both wholesale and instance-based moral criticism of emotions is that they don't usually focus on the second-personal aspects of morally problematic emotions. So while the target of an unfair emotion can criticize the subject on moral grounds, for example, that their unfair anger betrays a vicious care for social status, or that inapt contempt is corrosive and morally bankrupt, that type of criticism is just as available to anyone else.

To criticize someone's anger as expressing a problematic underlying character trait does not require making the criticism on behalf of anyone. When criticizing the angry queuer as being unfairly angry on behalf of the

teenager, the teenager could cancel out this criticism by saying that they are okay with it. It that case, you could no longer make the criticism on behalf of the teenager, without being patronizing, that is. You can, however, still criticize the queuer's anger as an expression of a vice or morally misguided in general. Hence, a criticism of unfairness has to capture this notion of directedness, which other forms of moral criticism do not.

2.5 Towards Unfairness Criticism

Throughout this chapter, I have shown that the more commonly discussed types of criticism of emotions differ from the notion of unfairness, which I have so far only tentatively outlined. On the one hand, neither criticisms of inherent incorrectness or unfittingness, imprudence, nor the irrationality of inconsistent emotions can capture the normative import of unfairness. General forms of moral criticism, on the other hand, cannot capture the directedness and conditionality of unfairness.

While the above types of moral criticism are indeed valid forms of moral evaluation of emotions, they don't amount to a criticism of unfairness. The type of criticism of an emotion that this book is about falls under the category of moral criticism. I agree with these types of moral criticism that the moral relevance of emotions lies in their power to motivate and influence our desires and behaviour, especially with regard to our treatment of other people. However, the type of criticism I am after differs from the aforementioned types of moral criticism in three important ways:

First, it is second-personal. This means that the emotion is criticized as a moral problem because of the person who is unduly targeted by the emotion. Additionally, the target is in a special position, that is, has standing to make this type of criticism, while others are not – or at least not directly. This differs from the aforementioned types of moral criticism where, in principle, anyone could criticize a morally problematic emotion, whether they are personally involved or not, even though there might be factors, like hypocrisy, that can undermine the credibility of someone making such criticism.

In contrast, second-personal criticism is limited from the beginning to those personally targeted by the morally problematic emotion. Other can at best make the same criticism on their behalf, but run the risk of overstepping and disregarding the target's moral autonomy. What I need to show, to argue this point, is that there is a significant burden emotions can put on specific other people and that they have a special criticism because of it.

Since, if emotions are merely something that affects the subject and are only contingently of relevance to others, it would be difficult to argue that we have any standing to interfere with other people's personal, internal life. To show this, I need to address what I call the *no-harm problem*, the

possibility that emotions, since they don't pose any direct threat to others, are not of moral concern to them. I discuss this problem in Section 3.1.

Second, the second-personal criticism I am after is less concerned with the evaluation of the subject's moral character or the general moral value of the emotion, and more so with the interpersonal attitudes the subject harbours towards the target of the emotion. Much of the second-personal nature of the sought criticism is due to it being based in the attitudes one person holds towards another. For example, if you are scorned by someone, it matters little whether they are generally a scornful person to how that scorn impacts your life. Or if someone is unfairly angry at you, you have an interest in them ceasing to be angry that is more urgent and goes beyond the question, whether anger has any general moral value.

Third, while it is not explicit in the accounts of moral criticism above what the normative upshot of these morally problematic emotions are, the criticism I am after should be open to the possibility of holding the subject in some way accountable for their unfair emotion. In the angry queuer example, something like an apology or expression of regret seems appropriate, and an account of unfair emotions should be able to make sense of that.

In the next chapter, Chapter 3, I present my own account of the unfairness-criticism of emotion. This account should be understood as an additional type of criticism of emotion, and not a rival interpretation to any of the above-discussed types. In Chapters 4 and 5, I discuss whether we can be accountable for unfair emotions. To that end, in Chapter 4, I first disambiguate different ways of holding people accountable for their emotions and what forms of sanction or even punishment are clearly unjustifiable. I also address what I call the *no-control problem*, namely, that we don't seem to have the same degree of control over our emotions as we have, for example, over our actions. However, control also seems to many philosophers to be a prerequisite for any form of accountability. I argue that we can overcome this apparent problem and present an alternative approach in Chapter 5.

Notes

1 It is important here to distinguish between a moral appropriateness standard for an emotion and an appropriateness standard for a moral emotion – meaning an emotion that involves moral considerations in its fittingness conditions. My interest here is in the former, not the latter.

2 See D'Arms and Jacobson, "The Moralistic Fallacy"; Rabinowicz and Rønnow-Rasmussen, "The Strike of the Demon"; de Sousa, *The Rationality of Emotion*; Deonna and Teroni, "Which Attitudes for the Fitting Attitude Analysis of Value?"; Bell, "Globalist Attitudes and the Fittingness Objection"; McHugh and Way, "Fittingness First"; Na'aman, "The Fitting Resolution of Anger."

3 For some prominent defences of judgement theories, see Nussbaum, "Emotions as Judgments of Value and Importance"; Solomon, "On Emotions as Judgments."
4 For some prominent defences of perceptual theories, see Prinz, *Gut Reactions*; Tappolet, *Emotions, Values, and Agency*; Döring, "Seeing What to Do."
5 Deonna and Teroni, "Emotions as Attitudes."
6 Deonna and Teroni, "Which Attitudes for the Fitting Attitude Analysis of Value?" propose a conception of *epistemic* fittingness that holds between any psychological mode or attitude that can be said to present the world in a certain way and its fitting-making conditions. Hence, it applies to emotions as well and not only to *traditional* epistemic mental states like beliefs, judgements, perception, or understanding.
7 See Deonna and Teroni, "Emotions as Attitudes," 19.
8 See, e.g. Deonna and Teroni, "From Justified Emotions to Justified Evaluative Judgements," 97, for a definition of justification that does not depend on a belief based picture of emotions.
9 Echeverri, "Emotional Justification," 545.
10 See Nussbaum, "Transitional Anger."
11 For such arguments see Rabinowicz and Rønnow-Rasmussen, "The Strike of the Demon"; D'Arms and Jacobson, "Sentiment and Value."
12 See, e.g. Döring, "What's Wrong with Recalcitrant Emotions? From Irrationality to Challenge of Agential Identity"; Brady, "The Irrationality of Recalcitrant Emotions"; D'Arms and Jacobson, "VIII. The Significance of Recalcitrant Emotion (or, Anti-Quasijudgmentalism)."
13 See Brady, "The Irrationality of Recalcitrant Emotions," 413.
14 Borrowed from D'Arms, "Value and the Regulation of the Sentiments."
15 Brady, "The Irrationality of Recalcitrant Emotions," 414–16.
16 Brady, "The Irrationality of Recalcitrant Emotions," 420.
17 Brady, "The Irrationality of Recalcitrant Emotions," 426 ff.
18 Under this interpretation, recalcitrant emotions would be bound by a *wide-scoped* rational requirement, cf. Broome, "Wide or Narrow Scope?"
19 See Silva, "The Epistemic Role of Outlaw Emotions," for a more detailed discussion of this example.
20 Bittner, "Is It Reasonable to Regret Things One Did?"
21 I disagree with Nussbaum's characterization of anger. I elaborate on some widely shared concerns about her account of anger in Section 6.1.2.
22 Nussbaum, "Transitional Anger."
23 While Nussbaum mainly considers the case of anger here, I take it that this applies to all emotions.
24 Bell, *Hard Feelings*, chap. 4.
25 See Hurka, *Virtue, Vice, and Value*; Adams, *A Theory of Virtue*.
26 Bell, *Hard Feelings*, 149.
27 Bell, *Hard Feelings*, 150.

3 Unfair Emotions

In the previous chapter, I have given an overview of various types of criticism of emotions that are commonly discussed in the philosophical literature. In this chapter, I present a type of criticism of emotions that has so far been given little attention. Namely, a second-personal moral type of criticism, which I will, somewhat stipulatively, call *unfairness*.[1] While I hope to capture some everyday associations of the term *unfairness*, I acknowledge that my account is not exhaustive of the many different use conditions of *unfair*, but that is not my primary intention, anyway.

As a point of disambiguation, I do not intend to capture the notion of distributive or comparative fairness. Comparative fairness in the context of emotions might be concerned with questions like: 'Is it not unfair if I am a calm and forgiving person with everybody else, even if it would be fitting to get angry, but then get – fittingly – angry with you?'; or 'Is it not unfair when I feel admiration for someone qua being a member of my own group that I don't feel for someone else who isn't part of the group?'[2] These examples are concerned with comparisons of how you respond differently to different people, and whether there can be injustices in doing so.

There might also be interesting questions about the distribution of emotional responses that are not necessarily comparative. For example, it might be the case that people require emotional attention to develop relevant social skills and that an unequal degree of emotional attention might lead to wider societal inequalities, but this is not what I have in mind here. While the term *fairness* has been widely used to denote questions of distribution – for example, in political philosophy and theory of justice, or in mathematical problems on procedures to divide cakes – it can just as well be used in the way I intend, for cases where one person treats another without due regard.

I am interested in a notion of *unfairness* towards a person in instances where someone is targeted by an emotion that is uncalled for. For example, when you have done nothing wrong and been respectful towards someone, and they still get angry with you for something innocuous, you might

DOI: 10.4324/9781003498032-3

feel treated unfairly. Other examples of the type of unfair emotions that I have in mind are being disappointed with your child for not living up to unreasonable expectations, feeling joy or glee at someone being harmed or hampered who does not deserve it, or being afraid of someone you perceive as threatening for no good reason, but rather out of prejudice. Although there are clear and worrisome systemic factors at play in some of these examples, my focus here lies on what happens, normatively speaking, on an interpersonal level – between the subject and the target of the emotion.

I think the distinction between comparative and non-comparative unfairness can also be seen in other contexts. For example, if I stand in a public space and freely distribute money to passers-by, that does not seem like an inherently unfair thing to do. However, if we add context, it can very well turn into a comparatively bad action. For example, if I only give out money to those who share my gender and skin colour, the comparative unfairness becomes immediately apparent. Of course, each instance of me giving money to a passer-by does not become non-comparatively unfair. The unfairness is extrinsic to those individual acts and pertains to the act only within the broader context. In contrast, if I go around stealing money from passers-by, each individual instance would constitute a non-comparative unfairness, regardless of whether I discriminate in my stealing behaviour or steal from everyone equally.

Comparative unfairness could be understood as treating people differently due to factors that should be irrelevant. For example, if you had (or have) two children who both smeared the wall with paint, and you sent one of them to their room and let the other off the hook, everything else being equal, that would be an unfair treatment of the grounded child. This is because all relevant factors that would warrant grounding are the same for both children, hence the differential treatment can only be based on some non-relevant factor – and therefore unfair. It would not be unfair if you let both of them off the hook or sent both of them to their room, but since you treat them differently due to arbitrary factors, grounding only one is comparatively unfair. While this kind of example makes it seem that comparative fairness is simple enough and can be understood completely independently of non-comparative unfairness, I still want to give one reason to think that they might be more closely related than initially obvious. Consider the case, where your children did nothing wrong. In this case, it would still be unfair to ground only one of them, but it does not seem more fair to ground them both. To the contrary, it would be *more* unfair to ground both children rather than just one, because now your treatment of both of them is unwarranted. There is, of course, a sense in which treating both children badly is comparatively more equal, but saying that it is therefore fair still seems wrong. It is as if there is no (comparative) fairness in overall unfairness.

One way in which we can resolve this unease by calling equal mistreatment 'comparatively fair' is to acknowledge that fairness is an asymmetric concept that includes non-comparative fairness as a necessary requirement. What I suggest is that those relevant factors which allow for differential treatment just are the factors that make any treatment of others non-comparatively fair or unfair. If one of your children has duped the other into smudging the walls and lied that it is allowed, then it seems fair to only ground the instigator but not the misled sibling. This is because there is now a relevant difference between the two: one of them knew they should not paint the walls, and the other one didn't know. Knowing that what you do is wrong is one of the features relevant to being grounded, or other forms of chastisement.

The opposite case works differently, however. Grounding one of your children while not grounding the other, when they both are culpable, is comparatively unfair, even though the grounding would by itself be warranted. This is because there are more factors that are relevant to comparative fairness than to non-comparative fairness. Non-comparatively, it is fair to ground any one of your children. Hence, grounding one remains non-comparatively fair, even if you don't ground the other. But in comparison, showing leniency to one but not to the other on no good grounds, is again comparatively unfair.

I don't intend to go too deeply into details about the conditions of comparative fairness. There is certainly an interesting discussion to be had about what features make it fair to show leniency to only one person but not the other. But my focus here is on the non-comparative features of fairness, as described earlier.

Unfairness, understood non-comparatively, seems to have three characteristic features that are relevant to understanding the question whether emotions can be unfair. These three features are meant to outline the notion of unfairness I have in mind. The focus of this chapter is to identify when and how an emotion could be the object of an interpersonal moral criticism that falls under this notion of *unfairness*.

First, the supposedly unfair emotion must have some morally objectionable aspect. This could involve aspects like harmful consequences, or a rights violation, or that the emotion instantiates a vice or is a product of a morally objectionable character. In Section 3.1, I argue that the morally objectionable feature of an unfair emotion lies in its motivational force and the type of behaviour it motivates. I argue that the various types of emotions inherently motivate specific types of behaviours, and that this motivational force can pose a moral hazard to others. This moral hazard can be morally evaluated similarly to other forms of immoral endangerment of others. However, not all morally objectionable emotions are unfair. The flavour of moral wrongness characteristic to unfairness has to be further differentiated.

The second feature of unfairness is that it is conditional. There seem to be circumstances under which, often due to the target's own prior behaviour, what would otherwise be considered unfair is not unfair under such conditions. For example, it seems generally unfair to not believe someone when they express how they feel. But when a chronic liar constantly deceives people to manipulate them, the liar has very much brought it on themselves that they are no longer trustworthy, and not believing them no longer seems unfair. Unfairness seems sensitive to context in this way. In Section 3.2, I argue that the unfairness of emotions seems to reliably co-vary with the emotion's fittingness conditions. I further propose an account that aims to capture why this should be the case. I argue that an emotion's fittingness conditions also provide *pro tanto* moral justification for the behaviour motivated by the emotion. This type of *pro tanto* justification is at the same time what the fairness of an emotion depends on.

The third feature of unfairness is that it has an interpersonal directedness. What this means is that there is always a person or group of people who is targeted by the moral wrong of the emotion. Rather than being concerned with evaluating someone's character, or the general good or bad that comes from an emotion, calling it unfair is primarily about the emotion's moral relevance to its particular targets. For example, rather than judging anger to be generally destructive or a sign of a vicious character, the question whether it is unfair would be concerned with whether there is someone who is wronged by the anger – in the intended sense, an emotion is not simply unfair, but always unfair *to* someone. In Section 3.3, I argue that for an emotion to be unfair, it has to be directed at the same target that is also the focus of the emotion's action tendencies. I further address the difficulty of conceptualizing this sense of directedness as a moral wronging. Namely, that to claim that an emotion can wrong others might require us to assume duties to feel emotions. Finally, I propose a solution that avoids such an assumption of duties to feel emotions.

3.1 Morally Objectionable

The first feature of unfairness, as I conceive of it, is that the unfair thing is in some way morally objectionable. When applied to emotions, this means that unfair emotions are in some way morally objectionable. This is also one feature that distinguishes the notion of unfairness from that of unfittingness, as discussed in Chapter 2. Criticizing an emotion as unfitting is at first morally neutral, while criticizing it as unfair always entails a moral criticism. For example, the fear of a deceptively snake-shaped stick which is not dangerous in any way is not by itself morally objectionable and does not necessarily reflect badly on the subject's moral character. In

contrast, the fear of a stranger based on a prejudice seems to be morally objectionable.

As mentioned in Chapter 1, such a comparison might initially trigger some scepticism about whether an emotion can really itself be morally objectionable. The mistaken fear of a snake-like stick and the mistaken fear of a stranger seem very similar in some respects. In both instances, the subject commits some type of error in how they perceive of or react to the target of their fear – they mistake the degree of danger or threat the target poses. It is not obvious what it is about one of these mistakes that would be morally objectionable, while the other one remains morally innocuous.

Moreover, there seem to be many phenomena closely related to the emotion of fear that are more obviously open to moral criticism – such as good or bad consequences, or the subject's virtuous or vicious character traits that lead to the emotion. When we look at an example like the prejudiced fear of a stranger and find it plausible that there is something morally wrong with that fear, a sceptic might object that we are misidentifying the source of our moral discomfort. It is not the fear itself that is morally objectionable, we are merely implicitly judging their character as morally deficient. Or when we assume that an impatient queuer's anger is unfair to the person in front, a sceptic might instead point to the effect the anger has on the other person as what is morally wrong. Such alternative explanations of what triggers our moral alarm-bells might lead us to conclude that it is never the emotion itself that is unfair, but only these related phenomena that give rise to the judgement that there is something morally wrong with prejudiced fear or unfair anger.

In this section, I address these two points of scepticism. First, in Section 3.1.1, I propose that what is morally objectionable about an unfair emotion is not simply that it misrepresents its target, but that it also poses a moral hazard to the target. This moral hazard can be criticized as morally objectionable, similar to unjustly endangering people.

Second, in Section 3.1.2, I argue that this moral hazard is not an accidental by-product of an unfair emotion, but that it is an inherent feature of the type of emotion in question. Given this inherent connection, it is the emotion itself and not any closely related phenomenon which is the object of our moral objection to an unfair emotion.

3.1.1 Moral Hazards of Emotions

In Chapter 2, I list several types of moral criticism of emotions that agree with my assumption that emotions can themselves be morally at fault. For one, Nussbaum criticizes anger not only on the basis that it misrepresents the facts about its target – although that may be part of the problem – but in the destructive potential inherent to that type of emotion. She identifies

the potential for harm in a desire for retribution that is inherently linked to anger.[3] This desire for revenge influences the subject's behaviour or patterns of thinking in a nefarious way, motivating them to harm or denigrate perceived wrongdoers in some way. In effect, the only thing achieved by anger's desire for revenge is to bring more pain and suffering into the world, instead of addressing or relieving the original wrong.

We don't need to agree with Nussbaum's entire criticism to see the merits of this approach. The focus on the potential for harm inherent to an emotion like anger is useful to illustrate the type of relevance an emotion can have on other people. While anger itself does not harm the target or affect their vital interests directly, it does produce an increased threat of the subject acting on a desire for revenge, which would harm or affect the target's basic interests.

One way in which Nussbaum's account does not sit well with the idea of unfair emotions is that it criticizes all instances of anger equally, which goes against the conditionality feature of unfairness. If anger is always morally objectionable, then I could not do anything to ever deserve someone getting angry with me. But it seems that if anger is sometimes unfair, then it also has to be fair at other times. I will say more on the conditionality feature in Section 3.2.

Another question is whether it is possible to generalize Nussbaum's case against anger to include other emotions. The desire for revenge seems particular to anger and is not present in, say, fearing a stranger based on a prejudice. Such a case of fear seems nonetheless similarly problematic as a case of unfair anger. In the following, I aim to develop a general account of unfair emotions that can account for emotions of any type that target other people and expose them to similar kind of threats.

To address this question, we need to establish what aspect of an emotion can even pose a morally problematic threat to others. To that end, we can draw on theoretical resources from related work in the ethics of beliefs. Cox and Levine formulate the idea that holding certain beliefs can be morally problematic in a similar way to the threats of anger. One way in which a belief can be immoral is by posing a *moral hazard* to others.[4] They define a moral hazard as a heightened likelihood of the subject acting immorally towards others. For example, if you believe that you can get away with stealing a large sum of money and not suffer any negative consequences, you will be more likely to do it. It might even be more rational to act in such a way, even if it were morally objectionable.

Just as beliefs can pose an increased likelihood of acting wrongfully, emotions like anger can as well. If anger does motivate vengeful behaviour, as Nussbaum claims, then you will be more likely to act in such a way that would harm someone or violate their vital interests. However, is an increased likelihood of wrongful behaviour already wrongful in

itself? I think this increased threat or hazard should be considered to affect the target's interests in a similar way to other forms of endangerment. The following example can help illustrate the analogy to other forms of endangerment:

Drunk Driver: The driver of a car gets drunk and drives home through a densely inhabited neighbourhood. If someone were to step out into the street, the driver would likely not be quick enough to avoid hitting them.

Even if nobody steps out into the street, the drunk driver's driving clearly constitutes a hazard to the people in the neighbourhood by increasing the likelihood of an accident. Creating and maintaining such an endangerment is in itself morally problematic, whether anyone actually gets hurt or not. While getting hurt or harmed would go against the very basic interest of bodily integrity, we also have a derived interest in safety from having this basic interest compromised, that is to say, an interest in *safety*.

While endangerment might not directly compromise the more basic interest in bodily integrity, it affects it indirectly by going against the derived interest in safety from bodily harm. Similarly, unfair anger might not directly interfere with any basic interest of its target, but it can affect it indirectly if it increases the likelihood of being treated aggressively that would compromise a basic interest. Put differently, unfair anger would go against a derived interest to not be in constant danger of being subjected to aggressive confrontations when you have done nothing wrong to deserve this.

There are several dis-analogies between the drunk driver case and emotions like anger, but the common feature I want to highlight is that an increased threat of harm can already be morally problematic. One difference between the drunk driver's driving and getting angry is the impression we have that the drunk driver has already acted in a certain way to get into this situation, but now no longer has any control over whether someone gets hurt once they step in front of the car. The driver's reflexes are simply too slow, and the distance it takes to brake too long. People getting angry, on the other hand, don't usually need to act in any specific way to become angry and, once they are angry, still have plenty of control over what they do while being angry. Because of this asymmetry of agency, it is plausible that the drunk driver violates a duty they have towards the inhabitants of the neighbourhood, while someone getting unfairly angry does not. This is because, as I discuss in Section 3.4, having a duty plausibly implies having at least some relevant degree of control over the object of your duty. The point here is not about the responsibility for unfair emotions or for driving drunk, but about what kind of moral status they have. I address the larger question of responsibility for emotions in Chapter 4.

In any case, the threat or hazard inherent to the two situations, which makes them morally problematic in the first place, is of a similar kind. It consists in an increased likelihood of the state in which the subject finds themselves to adversely affect some relevant or vital interest of other people. An emotion is only unfair if it increases the likelihood of the target being adversely impacted, like being harmed, being disadvantaged, or having their autonomy violated by the type of behaviour the emotion elicits. An increased likelihood of benefiting the target would not be considered unfair to the target, although it might lead to a form of comparative unfairness towards third parties.

However, simply an increase in likelihood of adverse behaviour towards others is not sufficient to call an emotion unfair, since there could be situations in which that behaviour is morally justified or even obligatory. For example, assuming that if someone tells you that something you did was horribly offensive to them, you should normally believe them.[5] It seems reasonable to assume that contempt leads you to, among other things, disregard someone's moral testimony. Hence, if you unfairly felt contempt for a person trying to give you this kind of moral insight, you would wrongfully be inclined to disregard it. In contrast, feeling contempt for a person because of their horrible moral opinions, and therefore being disinclined to believe them in this respect, would probably be the right thing to do. It seems, therefore, that the unfairness of an emotion is conditional on the fittingness of the emotion itself. I come back to this conditionality requirement of unfairness in Section 3.2.

Additionally, a simple increase in likelihood of morally problematic behaviour is too broad to capture only instances of unfair emotions and therefore not sufficient to differentiate unfairness from other forms of badness. There are several cases where an emotion would increase the likelihood of the subject acting wrongfully that don't seem to translate back to the emotion itself being unfair. The following examples illustrate two additional considerations to keep in mind.

First, if a subject is angry with their phone not working correctly and has an increased likelihood of aimlessly throwing it around, other people could get hurt by it. This increased likelihood does not seem to warrant the inference to the anger itself being unfair. Would the unfairness be towards the person more likely to get hit? That does not seem right, since the subject is not even angry at them, but at the phone. Is the anger therefore unfair towards the phone? Again, that does not seem right, the phone is not a moral patient who should be treated fairly. There seems to be a disconnect between whom the anger is about and who is affected by the elicited hazards. It seems that there should be some shared direction between whom the emotion is about and who is affected by the hazard. I discuss this feature of interpersonal directedness in Section 3.3.

Second, if a subject feels embarrassed whenever they meet or think of a specific friend because they have done something incredibly thoughtless in front of them, that might increase the likelihood of them forgetting to wish their friend a happy birthday. Again, this does not seem to warrant the inference back to embarrassment being unfair towards the friend. While an inclination towards aggressive or confrontational behaviour seems to be an inherent part of anger, forgetting things seems only arbitrarily connected to embarrassment. At best, embarrassment focuses the subject's mind on something else, leading them to not think of the birthday, but the same could happen with many other emotions, like anger or curiosity at something unrelated to the birthday. Producing a hazard is not sufficient to make an emotion unfair. What is missing is a clear link between the emotion's character and the motivation towards hazardous behaviour towards other people. I discuss this link in the following subsection.

3.1.2 Emotional Behaviour

So far, I have simply assumed that emotions can pose hazards to other people. But can such hazards really be attributed to an emotion, or are they mostly incidental by-products of emotional episodes – like forgetting someone's birthday because of embarrassment? What is required to establish a robust relation between emotions and hazards is to show that a specific type of emotion reliably disposes subjects to a specific type of behaviour. Put differently, it needs to be shown that there is such a thing as *emotional behaviour*, meaning behaviour that is both typical for a given emotion and motivated by the emotion. For example, it is not sufficient that getting angry with your dog one time made you forget the bread in the oven. While anger might have a motivational role to play in why you let the bread burn, it is a far too specific type of action to be reliably caused by anger.

Conversely, you might reliably be disposed to sneeze every time you are strongly amused by something. But it would be far-fetched to claim that sneezing is motivated by amusement, since it is closer to a reflex than a type of behaviour, and not really open to being influenced by motivation. Hence, in addition to being reliable, emotions need to play a central role in the explanation of a certain type of behaviour for it to count as emotional behaviour.

The type of behaviour that is more inherently produced and motivated by anger is a general tendency towards aggression or confrontation. This is a much more vague categorization of general behaviour, rather than specific actions, but it is also a more plausible description of how emotional behaviour is to be understood. There seems to be a general theme to what kind of behaviour an emotion would motivate, but the specific actions that would result out of it will be highly dependent on the situation. For

example, when afraid of a bear, you would feel motivated to protect yourself from it, but whether you run, hide or play dead will depend on what you think will protect you in this specific situation.

Are emotions reliably and inherently tied to specific types of behaviour? The answer depends to a large degree on how we understand the nature of emotions. However, I think that under most accounts of emotions in the philosophy of emotions, there is some understanding of emotional behaviour and of how emotions influence or motivate a type of behaviour that is typical to the type of emotion in question. The only theories of emotions that have some difficulty to establish such a link between emotions and behaviour are *bodily feelings* or *somatic* theories of emotions.

Early feelings theories, most famously proposed by William James, identify emotions with the bodily and hedonic feelings experienced during an emotional episode. If we accept such an approach, it is very unlikely that emotions play a reliable role in producing a typical kind of behaviour. James saw the direction of the association between emotions and action the other way around. He argues that it is more rational to assume that we feel sadness because we cry, and as an effect of crying, or afraid because we tremble, angry because we strike – rather than trembling as an effect of sadness or that we strike someone because we are angry.[6] The argument is that these bodily changes must come before the emotional experience because otherwise the emotions would not be felt, since what we feel during an emotional episode just are those bodily changes. However, most newer theories of emotion reject this order of explanation.

On the other end of the spectrum, motivational or attitudinal theories of emotions even draw a direct conceptual link between emotions and specific types of behaviour, identifying emotions with their motivational profile. Authors like Frijda, Scarantino, or Deonna and Teroni all associate specific action tendencies, or *action readiness*, with the different types of emotions,[7] going even so far as to identify emotions with their motivational profile.[8] For example, when in fear of a bear, a subject is inherently motivated to do certain things, like search for protection, run away or hide, that conform with the inherent motivational profile of fear. It is possible to identify such a motivational profile for a whole range of different emotions. Frijda lists anger as being associated with antagonistic behaviour, fear with avoidance, or interest with attending to its object.[9]

The more interesting question is whether judgement or perceptual theories can account for emotional behaviour. Judgement theories of emotion draw an analogy between emotions and judgements, in that an emotion is like an affirmation of an evaluation. As such, emotions would clearly be able to influence behaviour in the same way as any other judgement or belief can. Namely, emotions inform the subject about what feature of a situation is valuable, worth preserving, achieving, or dangerous and to be

avoided. For example, when you fear the bear in from of you, that fear is the judgement that the bear is dangerous and can exert some rational pressure, assuming that you value your bodily integrity, to form an intention to flee, or otherwise protect yourself from it. Perceptual theories can tell a similar story, in which fear is the perception of something as being dangerous. In this case, the rational pressure might be weaker and more easily defeasible by other judgements. But nonetheless remains a present inclination that has to be overridden by competing beliefs.

So both under judgement or perceptual theories, emotions can make certain courses of action seem reasonable or at least exert some rational pressure to act in accordance with what the associated evaluative property entails.[10] For example, when something or someone seems dangerous to you, it would be rational to protect yourself in some way from the potential harm. This means not exposing yourself to vulnerabilities towards others. But seeing this as the most rational course of action can be harmful when others genuinely depend on your help, for example, to get up again after a fall. By helping, you open yourself up to being vulnerable to them, for example, to being mugged if they only pretended to need help. But by fearing someone who is in genuine need of help, you would see yourself as justified in withholding the aid they require, thereby exposing them to further harm. Such a link between emotions and actions, as made possible by a judgement theory, is similar to the motivating force of any other evaluative judgement or perceptions. So, it is up to the perceptual or judgement theory of emotion to specify how closely linked the resulting behaviour is to the emotion.

But does this suffice to show that emotions necessarily dispose to certain types of behaviour that could potentially be hazardous? Even if there is some rational pressure to act in a certain way, it could still be the case that some instances of an emotion like anger are purely contemplative and lack any behavioural impact. This might be most clear in cases where you are angry with a dead person or afraid of something you know you cannot evade. In either case, any rational pressure towards a behaviour typical of anger or fear would be cancelled out by its impossibility. But these cases do not really undermine the idea that emotional behaviour can pose a moral hazard, but merely reduce the scope of potentially unfair emotions to those whose typical motivational effect is even a possibility. This reduction does not seem too high of a price to exclude anger with the dead or fear of the inevitable from the judgement of unfairness, if on the flip-side this still captures cases like the angry queuer as unfair.

More difficult cases are those, for example, in which wrongfully aggressive behaviour is still possible, but in which the person feeling anger does not experience any inclination towards acting on it whatsoever. This is mostly a possibility under a perceptual theory of emotion, where an

emotion is like a perception that involves an optical illusion. While it still represents its object in a certain way, you have other beliefs that completely cancel out its rational import. For example, you might feel angry at the person next to you in the queue for their annoying nervous foot-tapping, while at the same time be aware that you haven't eaten in a while and that normally, you would not consider this a behaviour worthy of getting angry over. In such cases, you might be able to completely cancel out the rational pressure towards emotional behaviour. However, it still remains the case that the emotion itself would have the same rational implications to favour a specific type of behaviour, were it not for additional, extrinsic factors, like your other beliefs, cancelling it out.

To accept that emotions can be unfair, we do not need to assume that emotions can fully be identified with their motivational profile, as attitudinal theories do. We only need to assume that the link between emotions and their associated type of behaviour is not an arbitrary one. Or more specifically, that a given type of emotion consistently increases the likelihood of a specific type of behaviour, either through an inherent motivational profile, through rational pressure, or even by its rational implications that would favour this behaviour if not cancelled out by other factors. If this is the case, having an emotion that consistently poses a moral hazard by motivating a type of behaviour in a given situation is itself a moral problem. Hence, both perceptual and judgement theories of emotions can account for the hazards of emotion I have been focusing on, just as well as attitudinal theory. The only type of theory that does not work well with my assumptions is a somatic theory along the line of James'. Since I cannot resolve the dispute on which theory of emotion is correct here, I am content with the fact that the majority of newer theories can account for genuine emotional behaviour.

The initial challenge was that, if behavioural tendencies are merely an accidental by-product of emotions and nor reliably produced by it, then any hazard posed by the emotion is itself accidental. As a consequence, the moral objection against the hazard could not be traced back to the emotion itself. In the case of the missed birthday call, it is not simply the embarrassment that led to the subject forgetting to call their friend, but very likely a combination of factors. It could only be ascribed to the emotion if embarrassment were to reliably motivate forgetfulness, which is likely not the case. But in cases such as the unfair fear, there is such a reliable relation between the emotion and a tendency to act in a certain way, therefore the hazard can be traced back to the emotion. Most theories of emotion account for such a reliable connection between emotions and specific motivational tendencies.

However, as stated earlier, a simple increase in the likelihood of behaving in a morally objectionable way does not always amount to the unfairness

of an emotion. Under certain conditions, there can be good reasons that speak in favour of the motivational force to behave in a way that would normally be morally objectionable. For example, when anger is a reaction to something that warrants a confrontational response, the hazard posed by its motivational force gains at least some normative support. In the next section, I examine this conditionality feature of unfair emotions.

3.2 Conditionality

The second feature of unfairness that I have outlined at the beginning of this chapter is that unfairness seems conditional on the fittingness of the emotion. What this means is that some things can be normally unfair but become fair under specific conditions. This is not an unusual feature for moral evaluations to have. But it is the type of condition under which emotions are fair or unfair that distinguishes the evaluation of an emotion as unfair from other types of moral criticism.

I argue, in Section 3.1, that emotions can affect the interests of other people by posing behavioural hazards to them. However, this does not amount to a moral problem, because affecting other people in this way is only sometimes problematic; namely, when it is undeserved – or as I will argue, when it is unfitting.

In this section, my aim is to show that the fittingness conditions of an emotion also partially determine the unfairness of an emotion. In addition to arguing for an extensional adequacy of fittingness, in Section 3.2.1, I aim to show how the fittingness conditions of an emotion are relevant for its moral evaluation as fair or unfair, in Section 3.2.3.

3.2.1 Extensional Adequacy

D'Arms and Jacobson famously argue that moral criticism and fittingness criticism of an emotion are distinct.[11] In their prime example, a joke can be funny and therefore amusement fitting, but at the same time, amusement can be morally problematic. This suggests it is possible that fitting instances of emotions are nonetheless morally problematic. But can they also be *unfair*? In this subsection, I want to show, with the help of a couple of examples that the non-comparative fairness and the fittingness of morally hazardous emotions seem to extensionally coincide. I will further argue in the subsequent subsections that this is not a mere coincidence, but that the unfittingness of a morally hazardous emotion is a necessary condition for its unfairness.

A first example that shows that an emotion's fairness seems to depend on its fittingness conditions is the case of unfairly fearing a stranger because of prejudice. One reason why we might think that you should not fear a

stranger is that the stranger has not given you any reason to be fearful, because he simply does not deserve to be feared. In this case, the target of the emotion has not done anything to give the subject any reason to feel the emotion. In contrast, there are certainly people who deserve others to be afraid of them. For example, if someone actually intends to hurt or harm you, fear would be both a fitting response and the threat posed to you also gives some moral justification for fearful behaviour, like avoiding or protecting yourself.

As a second example, imagine yourself as a child frequently getting teased by your brother and consequently getting angry with him. Whenever you get angry, you are more likely to yell at him. Thereby, your anger poses a hazard to him. On the one hand, imagine your strict mother telling you both that if your brother teased you again and made you yell again, she would send him to bed without dinner. You find this punishment overly harsh so next time he teases you, you find you have all–things–considered reason not to get angry, and thereby subject him to the consequences. In this situation, however, it would not be unfair to get angry. The moral objection against getting angry is not that it would be inherently inappropriate in a morally problematic way, but only that it would increase the probability of overall unjustified consequences. So it seems that what would by itself be a fair – and fitting – instance of anger is merely overruled by other moral reasons that don't bear on its fairness.

Imagine, on the other hand, you're always getting unfairly angry with your brother when he has done nothing to deserve it. Every time you yell at him without cause, your mother gives him a treat to make him feel better again. Let's assume this makes it a net benefit to your brother, and therefore, that you are increasing the likelihood of doing him a favour by getting unfairly angry. Despite the likelihood of an overall positive outcome, your anger does not become more fair. Its unfairness does not disappear simply due to the more likely overall good outcome. Rather, anger is unfair whenever it is also unfitting.

The final example, and the most popular one in the literature on fittingness, is that of amusement over the funny but morally problematic joke. Can we say that amusement at a funny joke or situation can be unfair to someone? While it can be morally problematic, I don't think it can be specifically *unfair*. As I have discussed in Section 3.1, amusement would only be unfair if it motivates some morally problematic type of behaviour, and if that hazard is directed at the target of the amusement, as I discuss in Section 3.3. These conditions primarily apply in cases where you are not only amused at a joke but amused by a person or group of people; and then only if amusement inherently motivates some sort of morally problematic treatment of the target. The best case for the hazards of amusement is probably in what behaviour it prevents rather than what behaviour it

motivates. If anything, amusement motivates a subject to not take someone seriously or to treat a grave matter as something light-hearted or playful.

For example, if you see someone falling down and find it amusing, this is most likely because you do not think that they were severely hurt. The behavioural tendencies of amusement might further motivate you to not treat the fall as grave, and not immediately rush to their aid. This might lead to a disadvantage or danger for the fallen person in the form of not receiving help when they need it. If you appraised the severity of the fall correctly, you would not have evaluated it as funny in the first place. Again, it seems like the funniness of the situation coincides with the acceptability of the motivational profile of amusement. There are certainly more subtle ways in which amusement can render people ignorant to the needs of other people. In such cases, one can argue that amusement can pose a moral hazard to people who are made fun of in a joke or who are the unwilling participants in a silly looking but actually rather serious accident.

Another example might be when someone is trying to bring up a serious matter, like someone's misbehaviour or the dangers of a course of action, and instead of being heard is ridiculed and made into a laughingstock. In such a case, two things are wrong with being amused: (1) it is not actually funny – in many such cases, there is no actual humour in what the person has to say; and (2) it is unfair to be amused at the person since this reaction poses a moral hazard to them – namely, an increased likelihood of not being believed or taken seriously.

In examples like these, it seems then that amusement is only unfair if it mistakes something that is grave or serious as something not be to taken seriously. This does not seem to apply to all cases of otherwise problematic amusement. Conversely, if you hear a funny joke at a funeral, being amused by it is clearly not unfair.

This is for two reasons. First, the amusement is not necessarily targeted at any person or group that would be wronged by it – assuming the main complainants would be the other mourners. I say more on this point in Section 3.3. Second, whatever makes the joke funny seems to give at least some reason to treat the subject of the joke as funny, playful and not take it too seriously. Whatever action tendencies are inherent to amusement, they don't pose a moral hazard to anyone at the funeral. Being amused might make it more difficult for you to keep a respectful silence or to sincerely express other emotions, such as your condolence. But these are considerations that don't bear on the fairness of the amusement itself.

These examples show that fairness is conditional and co-varies with the fittingness of the emotion – excluding, of course, cases of emotions that don't pose moral hazards at all. However, the connection seems puzzling at first glance: Why should fairness be connected to fittingness when I have gone to some length to show that they are different standards in

the last chapter? Can fittingness give us the criterion we are looking for? In Section 3.2.3, I formulate an account that aims to explain why the fittingness conditions seem to play a relevant role for the unfairness of an emotion. But first, I want to distinguish two kinds of problems which can lead to an emotion being unfair – namely, by being misplaced or being misguided.

3.2.2 Misplaced or Misguided Emotions

We have to be aware of how we define our fittingness standard. Especially since in many cases where the targeted person finds the subject's emotional response unfair, the subject might find the emotion perfectly appropriate. It can even be the case that both the target and the subject consider the emotion unfair, but can still find it intelligible why the subject had such an emotional reaction. For example, in the process of overcoming certain personal insecurities, Sam might still get angry too quickly at some perceived slights. Sam can recognize these reactions as unfair but still make sense of why he gets angry, perceiving the supposed slight as an attack on his social standing. Just like still being ashamed of his criminal father despite thinking it is not shameful, he can still care enough to make anger seem fitting.

In many cases, however, it is probably not completely clear whether a hazardous emotion is unfair or not. For example, Lee and Kim are siblings, and Kim is not on speaking terms with both of their parents. Lee has tried to mediate a conversation and Kim agreed to keep an open mind. But Kim did not listen and interpreted everything their parents said in an uncharitable way. After leaving their parent's house, Lee is disappointed with Kim. Realizing this, Kim accuses Lee of taking their parent's side and holds that Lee's disappointment is unfair.

Disappointment can certainly have negative behavioural consequences when it leads to Lee not supporting Kim to the same degree or with the same conviction any more. Lee withdrawing emotional support in such a situation when Kim needs it most can be hurtful, it is clearly directed at Kim and a non-arbitrary feature of disappointment. However, Lee can counter the accusation of unfairness by pointing out that while the disappointment might negatively affect Kim, Kim provoked such a response by not upholding the promise to keep an open mind. Disappointment is a fitting response to someone failing to meet an expected standard of conduct, and Kim has clearly failed that. Therefore, disappointment is a warranted emotional response from Lee's perspective.

In cases like this, it is the fittingness of the potentially unfair emotion that is at stake in the dispute between the parties. Both can agree that Lee's disappointment is inherently hazardous towards Kim. What they disagree about is whether Kim's behaviour is a fitting basis for Lee's disappointment.

How do we decide cases like this? In all these cases, it seems at first glance that the fairness of the emotions coincides with its fittingness. Kim deserves Lee being disappointed if Kim's behaviour was actually disappointing, your anger is fair if it is a fitting response to your brother's actions, and your fear of the stranger is fair if the stranger is actually a threat to you.

There are two types of cases we should distinguish. First, an emotion can be *misplaced*, meaning that it misconstrues the situation and unfair because of it. Second, an emotion can be *misguided*, meaning it is rationally warranted given the subject's values, cares, or attitudes, but can nonetheless be unfair to the target because those underlying values, cares, or attitudes are themselves morally objectionable. In either case, the emotion would not be deemed fitting by a well-informed observer. In the case of a misplaced emotion, the subject does not need to have any problematic attitudes that inform the emotion, but merely misinterpret some feature of the situation. They might have some false information about the target. For example, when your friend gets mad at you for arriving too late at the train stop, not knowing that you were delayed by an unforeseeable traffic accident, their anger is misplaced and therefore unfair, but warranted from their perspective.[12]

The second type of situation might come closer to the joke example. If the amused subject finds the joke funny because of its sharp wit or absurdity, there might be little to worry about. We can see the humour even in jokes about serious topics, without taking those topics lightly in response. However, if the joke is of the sort that in order to find it funny, one needs to hold certain morally problematic attitudes in the first place, the amusement might well be unfair. This might be clearest in cases of *schadenfreude*. If you find it funny that your colleague, who always seems so larger-than-life and perfect, trips and falls, and now walks around with a limp, we can safely consider that unfair amusement. I'm assuming here that the hypothetical *you* in this case draws their amusement from an attitude of envy and malice towards their colleague, and not from simply not realizing the severity of the situation.

The two notions of misplacement and misguidance are thus more useful in identifying unfair emotions than unfittingness. That is not to say that they don't overlap in many cases. Emotions can be unfitting because they are misplaced or because they are misguided. So, in the case of Kim and Lee, their difference of opinion on whether Lee's disappointment is unfair probably depends on either Kim thinking it is misplaced, that Kim did not fail to meet a reasonable degree of open-mindedness, or that it is misguided, that Lee's expectations were unreasonable in the first place. If my account is correct, disagreement about the fairness of a directed hazardous emotion will always involve disagreement about another matter, either features of the situation or about the viciousness of certain underlying attitudes.

3.2.3 Intrinsic Moral Justification

While an emotion being fitting is not the same thing as it being morally right, there seems to be a close connection between an emotion's fittingness and its fairness. In part, this is possible because fairness is not an all–things–considered evaluation, but can be outweighed by other moral considerations. However, not all unfitting emotions are also unfair, since they may often lack the other features of unfairness, and be morally neutral or not directed.

I propose the following necessary conditions to capture the connection between fittingness and fairness: An emotion is *fair* if its fit-making conditions also provide *pro tanto* moral justification for the type of behaviour motivated by the emotion's inherent action tendencies. On the flip-side, an emotion is *unfair* only if the emotion is unfitting and thereby also fails to provide the fit-making properties that would have rendered it *pro tanto* morally justified in a way specific to the type of emotion in question.

The following example should illustrate the point: I will assume that, absent any justifying reasons, you should not behave in an aggressive or confrontational manner towards other people, and that anger towards someone motivates just such a type of behaviour towards the target. If the target of your anger actually did wrong you in some way, that injustice would make anger fitting. In addition to making anger fitting, the injustice also, other things being equal, provides a *pro tanto* justifying reason to behave in a confrontational manner towards the target. Hence, the fit-making property of anger also provides a reason for anger's action tendencies. Feeling anger, and with it the motivation to act in a confrontational manner, then simply means reacting to the fit-maker of anger – the injustice – with an appropriate mental stance.

This very specific type of *pro tanto* justification depends on reasons that intrinsically justify mental state. But while intrinsic reasons typically only justify a mental state in an epistemic or rational sense, here they also provide a limited moral support for the emotion. Therefore, we can speak of an *intrinsic moral justification*. This justification remains *pro tanto* because there might be other, overriding reasons that make it all–things–considered wrong to behave in a confrontational manner. For example, because confrontation would only escalate the situation into a much worse one. This might be an overriding reason to not act on your anger. However, it would neither render your anger unfitting nor would it make it unfair. This is because the fit-making conditions for your anger would still give it some degree of moral justification to its action tendencies, namely, that having been wronged remains a valid reason to confront the wrongdoer.

If the perceived injustice did not actually happen or does not actually constitute an injustice, then anger is not fitting and there would also be no

intrinsic moral justifying reason to behave confrontationally. There is no fit-making property to provide any such intrinsic moral justifying reason for the confrontational behaviour. Hence, absent the evaluative property that would make anger fitting, angry behaviour is not intrinsically morally justified. There might, again, be external factors that provide some reasons to be angry or even render angry behaviour all–things–considered the right thing to do. But if these other factors do not come from the fit-making property of anger, the emotion remains unjustified in this emotion-specific way. Put differently, it is possible that reacting with unfair anger would be the best thing to do, all–things–considered. For example, it might be overall good to get genuinely angry with an innocent person for supposedly littering if it leads to many other people to more diligently dispose of their rubbish. However, that does not mean that it was fair, since the justifying reason comes from factors external to the fittingness conditions of anger.

This explanation can account for all the aforementioned examples. To start with unfair amusement, I defend that whatever aspect of a joke or situation can reasonably be said to be funny or amusing provides an intrinsic moral reason to treat it as funny – meaning whatever behaviour is inherently motivated by amusement. If the action tendencies of amusement involve not taking something seriously, then this would amount to being intrinsically morally justified in taking that specific aspect of the situation not seriously. If there are other action tendencies for amusement, something like playful engagement, then you would be similarly intrinsically morally justified to playfully engage with that specific aspect. However, if you are amused at an aspect that cannot reasonably be considered funny, like someone drawing attention to a serious matter, then any such intrinsic moral justification is absent.

The same reasoning holds for the case of unfair fear. Avoiding or behaving defensively towards someone gains intrinsic moral justification if the person in question actually poses a threat – the fit-making condition of fear. Fear is therefore fair only if it is also a fitting response, namely, a response to the fit-making threat. This source of intrinsic moral justification is not given if the stranger is not threatening you in any way – if fear is not fitting.

And finally, confrontational behaviour is intrinsically morally justified if someone has acted offensively towards you, and those are exactly the conditions under which anger is fitting. Conversely, if someone has not done anything wrong or offensive, then acting confrontationally, as if they did something wrong, is not intrinsically morally justified in that way. Hence, anger is not fitting, and its action tendencies are not intrinsically morally supported due to the lack of its fit-making conditions.

3.3 Interpersonal Directedness

The third feature of unfairness is that it is inherently interpersonal. For example, if you get angry about something on the internet and in a fit of rage throw your phone across the room, there is an increased likelihood of other people in your vicinity getting hit and hurt by your flying phone. However, in such a case, your anger is not specifically unfair to those people because your anger is not about them. The person in danger of being hit by your phone is simply an unlucky bystander. It might be true that, to the person in danger of being hit, it makes little difference why they were exposed to such a hazard. But it makes a difference to whether the hazard posed makes the emotion unfair or not.

First, in Section 3.3.1, I aim to clarify in what sense an emotion can be targeted at a person in the sense relevant for unfairness. Second, in Section 3.4, I elaborate on how the directedness of the emotion corresponds to the directedness of the moral wrong entailed by an unfair emotion.

3.3.1 Targeted Emotions

Compare the above example of throwing your phone across the room to one where you are angry with your friend and therefore more likely to not help them move or support them when they are in distress. Despite being less severe than a flying phone, these actions motivated by your anger can pose a hazard to your friend. In this case, the person affected by the hazard posed by the emotion and the target of the emotion are the same person. The increased likelihood of negative consequences for your friend is not a random side effect of getting mad about something else, but the main behavioural impetus of your anger towards them.

What this example is meant to illustrate is that there is a special class of motivated action or behavioural tendencies inherent in an emotion that are specifically directed towards the target of the emotion. The behaviour an emotion motivates is often not aimless but targeted at a particular object. Anger at a friend does not merely motivate confrontational behaviour, but confronting *your friend*; and fear of the stranger does not merely motivate avoidance or hiding, but hiding from or avoiding *the stranger*. In cases where this behavioural tendency poses a hazard, the hazard is also directed at the target of the emotion.

The type of non-comparative unfairness introduced at the beginning of this chapter has the feature that it is always directed at a specific person or group of people. Emotions can have this feature if the hazard posed by their inherent behavioural tendencies is directed at a specific person or group of people. More specifically, emotions pose such a directed hazard if they are about someone, like fear of the stranger, and at the same time motivate a type of behaviour toward that same person, like avoidance of

the stranger. In such cases, we can not only say that an emotional episode poses some hazard to a person, but that the emotion is unfair *towards* the targeted person. This rules out people affected by fits of anger where the anger itself is unrelated to them. It also differentiates hazardous emotions in general from emotions that are unfair to specific people.

However, this way of tying the directedness of the hazard an emotion poses to its target leads to the next difficulty. Namely, in order to establish when an emotion is unfair to a specific person, we have to establish under what conditions an emotion is actually about the person.

3.3.2 The Objects of Emotions

Deonna and Teroni[13] suggest that the directedness of an emotion can be described in at least two different ways, either by describing the object nominally or propositionally.[14] An emotion can be described as directed at a nominal object, for example, when you are afraid *of the dog*. In this case, your emotion is first and foremost about the dog itself and not any particular aspect of the dog. However, your fear could also be described as directed at a propositional object, for example, you might be afraid *that the dog will bite you*. In that case, the emotion is directed at a specific state of the world, namely, one in which the dog bites you.

However, it seems that when we describe you as being afraid of the dog, we could ask for more clarification, for example, what about the dog you fear, and find out that you are actually afraid of the dog biting you. This seems true for many emotions that can be described as directed at a nominal object. You are not simply angry at your friend, but also that your friend broke your pen; and you are not simply proud of your daughter, but also that she won the football game. It is possible that whenever we sufficiently clarify what it is we actually fear, it will always no longer be the dog as a whole, which we would describe in a nominal form, but always something more specific, which can be described in propositional form. This raises the worry that we will also never have the case where a person is the actual target of an emotion. If no emotion ever actually targets a person or group of people, then it cannot be directed in the way I have been suggesting, and therefore emotions cannot have the directedness feature of unfairness – hence emotion cannot be unfair.

Is this really a problem, or is there a sense in which we can reformulate the description of an emotion as being directed at a propositional object back into one with a nominal object? Take, for example, the case where I'm ashamed *of being seen with you*. This emotion could just as well be described as me being ashamed *of you*, and *vice versa*, being ashamed *of you* might just as well be described as being ashamed *of being associated with you*. If these are both valid ways of describing the same emotion,

then we could identify emotions that are possibly unfair towards a specific person by the set of emotions that can be described by using propositional objects that can just as plausibly be described with the person as their nominal object.

For example, if 'I am afraid that you might hurt me' can just as well be expressed as 'I'm afraid of you', then we could say that the fear of what you might do has you as its target. It is certainly not the case that in such an instance, where I am afraid that you might hurt me, that I am afraid of everything about you. However, it is not very plausible that this is ever the case when the sentence 'I am afraid of you' applies, either. The reverse seems to be more problematic. Since being afraid *of you* is too general, its nominal object, *you*, cannot be easily translated into a propositional object, like being afraid *that you might hit me*, because the fear could be related to something else, like being afraid that you could interfere with my goals in other ways.

It is also not the case that any propositional object which involves a person can plausibly be reformulated using a nominal form. For example, 'I am angry that you lost the match' does not necessarily describe the same emotion as 'I am angry with you'. It could be the case that I am angry with the referees for not giving you the decisive point. Also, 'I admire your lack of shame' might not necessarily describe the same emotions as 'I admire you'. However, as long as it seems adequate to reformulate descriptions with a propositional object into a nominal form, it would be appropriate to take the emotion as fulfilling the directedness requirement of unfairness.

So, being angry *that you lost the match* would probably not be unfair if it cannot also be described as being angry *with you*. But, being afraid *that you might hurt me* could be unfair, given the fear poses an unjustified hazard, since it can plausibly be described as being afraid *of you*.

3.3.3 The Targets of Action Tendencies

Besides being about a specific person, we also need to establish in what sense an unfair emotion targets that person with its behavioural tendencies. Ronald de Sousa lists several ways in which emotions are directed at things.[15] For one, he distinguishes between an emotion's *particular* object – the situation or concrete object the emotion is about – and its *formal* object – the evaluative property of the particular object that the emotion highlights. Additionally, he proposes that emotions have certain aims that are characteristic for the type of emotion. For example, just as fear generally has the formal object of danger, it generally has the aim of establishing or regaining safety or protection from that danger. I suggest that a general aim receives a more specific focus when the emotion type is instantiated. A specific instance of fear would be the fear of a dog, where

the dog is the particular object of that emotion episode, which is evaluated as dangerous. In the same way, the aim of the emotion episode becomes to establish safety or protection from that dog, and not to establish safety in general. Hence, the dog is both the object of evaluation and the focus, or target, of the motivational aims of that instance of fear.

Can an emotion's object and the target of its action tendencies be different? The standard cases, like being afraid of the dog or angry with your friend, don't seem to support such a possibility. It is hard to imagine a situation where you are afraid of one thing, but that fear primarily motivates you to avoid or protect yourself from something completely different. We might think of an example, where you are afraid of accidentally meeting your boss when you are pretending to be ill, and therefore you avoid the part of town in which your workplace is located. We would not say that you are afraid of that part of town, so the object of your fear and that target of your action tendency seem to diverge. But, of course, you only avoid that part of town because you try to avoid your boss. There is nothing about the part of town other than its proximity to where your boss is likely to be, which makes it something you want to avoid.

There might be other emotions that are more promising when it comes to their objects differing from the targets of their action tendencies. For example, when grieving for a dead friend, the object of your emotion is either your friend or something like the things that connected you and the person they were. However, the action tendencies of grief, pulling back from activity, being more passive and contemplative, don't seem to target anything in specific. This might, on the one hand, be a misrepresentation of the action tendencies of grief, and we might rather understand them as very much directed towards contemplation of your loss. On the other hand, emotions like this, which don't have clear targeted action tendencies, are also not prime candidates for being unfair towards anyone.

To summarize, emotions are unfair to a specific person if two ways in which they are directed fall together. First, the emotion has to be directed at a person as its nominal object. This means the emotion is centrally about the target, for example, when you are afraid of someone but not of something specific they might do. Second, the emotion's motivational force has to be focused on the same target, which the emotion is *about*. This means the moral hazard posed by the emotion's behavioural tendencies is directed towards the target. This is the case, for example, when you fear a person and feel the urge to defend yourself or hide from them.

3.4 Directed Wrongs

So far, I have outlined the conditions under which the hazard of an emotion would be targeted at a specific other person. Namely, when

the emotions are about a specific other person and motivate a specific morally objectionable treatment of that person. These conditions nicely capture cases of emotions that are unfair to their targets. For example, these conditions manage to separate unfair anger from otherwise morally problematic cases of anger; or when disappointment is unfair because it is directed at a person but not when it's simply about something involving the person. However, the conditions discussed so far only establish that these kinds of directed and morally hazardous emotions can be morally objectionable because they endanger their targets, but not that they actually *wrong* their targets – that the wrong posed by the emotion is directed.

In a certain sense, what I have argued so far could already sufficiently establish the directedness of an emotion's unfairness. There is a minimal sense of directedness that is already satisfied by establishing that an emotion can be morally wrong, and that this wrong affects certain specific people that are at the same time the target of the emotion. However, there is a difference between the claim that something morally objectionable negatively affects someone and that the moral wrongness is itself directed. To claim that an emotion is unfair *to* its target seems to imply the latter, that the moral wrong itself is directed, or, put differently, that the emotion *wrongs* its target. This notion of *wronging* is typically used in a specific way in the philosophical literature that goes beyond the minimal sense of directedness, of a morally objectionable thing that negatively affects specific people, and implies the directedness of the moral wrong itself. Usually, this inherent directedness of a wronging is explained by the fact that someone's moral claims or rights are being violated. But the idea that there might be a moral claim to other people's emotional states seems quite implausible. This poses a substantial challenge to the idea that emotions can be unfair. In the following, I first expand on this challenge and subsequently argue that there is a different account of wronging which can nonetheless apply to emotions just as well as to actions.

3.4.1 Wronging and Claims

The type of moral wrong that seems inherent to the criticism of unfairness is what is commonly called a *wronging*. The idea of wronging is closely tied to so-called second personal moral theories, that put people's rights, claims, or interests at the centre of morality.[16] A central claim of second-personal moral theories is that actions that morally wrong another person are wrong because they violate a person's moral claim. Some second-personal ethical theories might go so far as to say that only matters concerned with people's claims are morally relevant,[17] but we don't need to go that far to use the notion of second-personal wrong to understand unfairness. It is sufficient

to acknowledge that a second-personal wrong, or *wronging*, is a distinct kind of moral wrong, which is at issue here.

In the case of a wronging, the wronged person has a special type of complaint against the wrongful act in question because it violates a moral claim of theirs. This fits nicely with the idea that unfairness involves a special type of complaint or criticism which is distinct from the criticism of incorrectness or even the other types of moral criticism discussed in Chapter 2. Specifically, it is a distinct type of wrong in that it is directed towards another person. To feel an emotion that is unfair to someone implies that the target has such a special complaint against the subject feeling the emotion. Any third person making the complaint would need to make it on the target's behalf.

The idea that to wrong someone means violating or infringing on their moral claims is an elegant way of analysing the directedness of a wronging. But, this identification of wronging with claim-violations raises some problems for my account of unfair emotions. It is a common principle, endorsed by most authors writing about wronging as claim-violations, that there is a close correlation between claims and duties.[18] The principle states that whenever A has a claim towards B that B φ, then B has a correlative duty towards A to φ. So if we were to assume that the target of an unfair emotion has a claim against the subject to not feel that emotion, then it would follow that the subject has a duty to not feel it.

Hence, this analysis of wronging requires that if we hold that emotions can wrong other people, we also need to assume that we have claims to or against other people feeling certain emotions towards us. This, in turn, implies correlative duties of the subjects to feel or not feel those emotions. The main worry that arises from the assumption of duties to feel is that we don't seem to be able to feel at will, and therefore could often not fulfil such a duty. However, a systematic inability to conform to duty violates the broadly acknowledged principle that *ought implies can*. Namely, that we cannot have a duty to do something that is impossible for us to do.[19] For example, if you cannot swim and are in danger of drowning, due to no fault of your own, you would plausibly have a claim against anyone to safe you, if doing so were of little cost or danger to themselves. If I were close by and a trained lifeguard, I would have a corresponding duty to hop in and try to save you. If I did not do so, I would wrong you by violating your claim against me. In contrast, if I fell out of the sky with you standing on the ground, lacking the ability to save me – since you can't fly – I would not have any claim against you to do so, and you would not wrong me by not flying up and rescuing me.

In the case of emotions, this seems implausible since we lack the necessary kind of voluntary control over them. We cannot simply start or stop feeling an emotion at will, whenever someone has a valid claim to or

against us feeling it. But if it is not the case that we can comply with a corresponding duty to feel, then we cannot really have such a duty. If it can't be assumed that we have such duties to feel or not feel emotions, then we cannot explain the wronging that unfair emotions seem to entail with the violation of a claim. But it nonetheless seems like an unfair emotion does wrong its target. Is there any other explanation of how an unfair emotion could constitute a wronging, or are we left with the conclusion that unfair emotions don't ever actually *wrong* their targets in the sense introduced at the beginning of this chapter? In the following, I explore an alternative explanation that does not base wronging on claims or duties, but on the possibility of being accountable for it.

3.4.2 Wronging and Accountability

While the idea that a *wronging* simply amounts to the violation of someone's claim is widely accepted, there are some cases in which who is wronged and whose claim is violated seem to come apart. Specifically cases, where the person being wronged, and to whom one is accountable or to whom apology is owed, is not the same as the one that holds the violated claim or right. Cornell, for example, argues that this coming apart can be seen in the case of actions that violate someone's rights but also thereby wrong third parties who do not have such a right but are affected by the action.[20] Such cases call for an alternative explanation for how we should identify who is being wronged by an action, namely, by identifying to whom the agents become *accountable* after having performed the wrongful action.

This idea, that when having wronged someone, you also become accountable or answerable to them, is a common second feature of directed or second-personal wrongs.[21] That is, not only are you generally responsible for the wrongful action and open to being blamed for it, but rather you are open to being blamed by the wronged person specifically, or by others on the behalf of the wronged person. Since the wrong of the action is directed, your accountability, and any duties of reparation or apology are also directed, in the sense of it being owed to the wronged person. In this context, we can distinguish two aspects of morality that are often intermingled: the *ex ante* and the *ex post* justification of and accountability for actions. On the *ex ante* side, before we act, we deliberate about what to do by invoking such notions as claims and duties. On the *ex post* side, after the fact, is where questions of accountability, damages or who has been affected or wronged by an action are raised.

This allows us to think of the direction of wrongfulness, not as a feature of *ex ante* morality, where it is defined by claims and duties, but as a feature of *ex post* morality, where it is identified by questions of blame and accountability. Building on Cornell's suggestion, my proposal for how to

identify a directed wrong is the following: On the *ex ante* side, if A has a *right* against B to not X, then B has a duty to A to not X. On the *ex post* side, if B *wronged* A with X, then B is accountable to A for X. That is not to deny that in most cases where B has neglected a duty to A, it is also the case that B has wronged A. I do not think that the exact same reasoning that applies to third-party wronging also holds for the case of unfair emotions, since third parties can be affected by emotions that are nonetheless not unfair to them. The main point here is that we can distinguish the concepts of rights and duties from directed wrongs. While this distinction separates the question of who has been wronged from the question of whose claims have been violated, it does not completely divorce the two. It is still a possibility that claims play a relevant role in determining who has been wronged. However, it opens up a conceptual space where it is possible for a wronging to not necessarily be based on the violation of a claim. I am also not claiming that B can wrong A with X despite being morally permitted to X, but rather that the categories of duty and permission don't apply to emotions, but the category of wronging does. Therefore, the category of permission does not apply here. So, rather than B wronging A despite being permitted to feel an unfair emotion, B can wrong A with an unfair emotion, without violating a duty against feeling the emotion, but also without being permitted to feel the emotion.

The alternative approach proposed here is to identify a wronging as a wrong that involves a specific type of normative upshot. More specifically, we can identify a wronging because it is a wrong that grants a specific person or groups of people special standing to blame, to forgive, or whom the perpetrator owes a justification – in a word, to hold the subject *accountable*. In addition, I suggest that being accountable in this way does not necessarily require any *ex ante* claims or duties. This does not mean that the person wronged is the only person who can blame the offender, but that, as mentioned, third-personal blame is always in a sense made on the wronged person's behalf.

Following this approach, the question whether unfair emotions can constitute directed wrongs turns on whether we can ever be accountable for something to which we never had an *ex ante* duty. If accountability does not require any violations of duty, then what else are the necessary conditions? And does an emotion ever give its target special grounds or special standing to hold the subject accountable? I discuss these questions – whether we require a certain amount of control to be accountable and whether the target of an unfair emotion ever has special standing to hold the subject accountable – in the remaining chapters. In Chapter 4, I argue that there are many ways in which we can *hold accountable*. I argue further that there is at least one way of holding people accountable that both accounts for this directedness feature of unfairness and applies to emotion – namely, by way of the reactive attitudes.

In Chapter 5, I show that the practice of holding people accountable can apply to emotions just as well as to actions. Holding someone accountable for their emotions entails holding them to an expectation of having an attitude of basic human regard towards the target; and to be disposed to feel certain reactive emotions in response to an emotion that reflects a violation of this expectation. Hence, an emotion wrongs the person who would have special standing to blame or forgive, if the emotion reflected an attitude of disregard towards them.

Notes

1 It might also be suitable to use a term like *undue* for this type of criticism, for its connotation of something being *due* to someone. However, undue can be read more easily as a non-moral criticism, and unfairness is simply a more captivating term. Second, while I don't intend to give an exhaustive analysis of the term *unfairness*, I don't think that *unfairness* has a single meaning that covers all possible use-cases, in any case, and at least one way of using it nicely fits my purposes here.
2 Many thanks to Daniel Telech, Laura Silva, Rachel Achs, Macalester Bell, and Justin D'Arms for these examples and drawing my attention to this distinction.
3 See Nussbaum, "Transitional Anger," 42.
4 Cox and Levine, "Believing Badly," 216–21.
5 The question whether one should ever accept moral testimony is disputed among philosophers, I only employ it as a point of illustration here. See Sliwa, "In Defense of Moral Testimony," for a defence of the claim; and Hills, "Moral Testimony and Moral Epistemology," for a negative case.
6 James, "What Is an Emotion?," 190.
7 Frijda, *The Emotions*; Scarantino, "Do Emotions Cause Actions, and If So How?"; Deonna and Teroni, "Emotions as Attitudes."
8 Scarantino, "The Motivational Theory of Emotions," 168.
9 Frijda, *The Emotions*, 88.
10 See, e.g. Brady, "The Irrationality of Recalcitrant Emotions"; Döring, "Seeing What to Do."
11 D'Arms and Jacobson, "Sentiment and Value."
12 One could argue that this is Nussbaum's first criticism of anger mentioned in the previous chapter; namely, that anger is misplaced if in response to a wrong. The subject erroneously assumes that anger can remedy or address the wrong, but in reality it can't and such assumptions are nothing but *magical thinking*; see Nussbaum, "Transitional Anger," 47–48. However, I am not primarily focused here on whether this is a good interpretation of Nussbaum's argument.
13 Deonna and Teroni, *The Emotions*.
14 I am limiting the discussion here to descriptions of emotions since not all theories of emotion ascribe content – whether propositional or nominal – to the emotions themselves.
15 De Sousa, *The Rationality of Emotion*, chap. 5.
16 See, e.g. Darwall, "Respect and the Second-Person Standpoint"; Darwall, *The Second-Person Standpoint*; Thompson, "What Is It to Wrong Someone? A Puzzle about Justice"; Wallace, *The Moral Nexus*; however, Owens, *Shaping the Normative Landscape*, defends a version of second-personal moral theory that

allows for so-called "bare wrongings", like breaching a promissory obligation, that don't violate any interests.

17 See, e.g. Wallace, *The Moral Nexus*.

18 For some examples, see Hohfeld, "Some Fundamental Legal Conceptions as Applied in Judicial Reasoning," 31 ff.; Wallace, *The Moral Nexus*, 2 ff.; Thomson, *The Realm of Rights*, 40 ff.; Feinberg and Narveson, "The Nature and Value of Rights," 243 ff.

19 I assume here that it is not possible to have a duty that one ought not to comply with. This is the subject of a larger philosophical debate, which I won't go into here. I merely assume, for the sake of argument, that if it is systematically impossible to comply with a certain duty, then we cannot assume that it is actually binding.

20 Cornell, "Wrongs, Rights, and Third Parties," 115 ff.

21 Darwall, *The Second-Person Standpoint*, 7–10.

4 Responsibility and Control

In the last chapter, I have argued for a distinct type of moral criticism that applies to emotions, namely, *unfairness*. An unfair emotion poses a directed wrong to the person or people targeted by the emotion by exposing them to a moral hazard – an increased likelihood of the subject treating the target wrongfully. I have claimed that the directedness of the wrong involved in an unfair emotion is indicated by the target having a special complaint against the subject or a special standing to blame. This implies that unfair emotions are, at least in principle, a valid basis for the target to blame the subject. Put the other way around, emotions can only be unfair if, under a certain set of circumstances, the subject can justly be blamed by the target for an unfair emotion.

This is not to say that a specific emotion is only unfair if the subject is blameworthy for it. Sometimes, the subject is not to blame, even if their emotion is an unfair one. There can be excusing circumstances or justifying factors that interfere with the blameworthiness of a subject in a given situation. For example, if the unfair emotion is based on a faultless misunderstanding or factual mistake, what I call a *misplaced* emotion in Section 3.2.2, it might still be unfair to the target, but the subject might nonetheless be blameless. Rather, what I claim is that if emotions were something that we could categorically not be blamed for, then it would be a mistake to call any emotion unfair. Just like, if actions were something you could not be blamed for, then no action would be a wronging – they could merely be bad or undesirable. If unfair emotions are something we can sometimes be blamed for, I need to show that, at least in those cases, we are responsible for them. I will not go too deeply into why it needs to be the case that we can be responsible for emotions in order to be appropriately blamed for them. I take it that blame is a form of holding someone accountable for a wrong for which they are morally responsible.

In Section 4.1, I first outline the argument against the claim that we are ever to blame for an emotion and the different ways this argument could be rejected. My own proposal is to disambiguate different notions of blame

DOI: 10.4324/9781003498032-4

and directly examine whether they require some type of control on the part of the subject, and what kind of control that would be.

In Section 4.3, I present different types of control that could be relevant for the justification of blame and whether we can reasonably assume that we have this type of control with respect to our emotions. I distinguish between direct and indirect forms of *voluntary* control and a type of *rational* control that is commonly discussed in the literature on responsibility for mental states.

In Section 4.4, I propose different ways of holding someone accountable that illustrate possible interpretations of blame – *attribution*, *answerability*, *social sanctions*, and *reactive attitudes*. The type of control required to justify holding someone accountable according to these interpretations varies in informative ways. Hence, not all these ways of holding a subject accountable are justifiable responses to an unfair emotion. I argue in Section 4.4.4 for the merits of the reactive attitudes (RA) theory – more specifically, that *resentment* is an appropriate conception of blame. I continue my argument that emotions are sometimes the proper target of resentment in Chapters 5 and 6.

4.1 The No Control Problem

My main thesis is that we can sometimes be blamed for our unfair emotions towards other people. A big challenge this thesis faces is the prevalent assumption that control is one of the necessary conditions for being responsible for anything,[1] combined with the common intuition that we cannot control our emotions as we can control our actions. If we cannot control our emotions, then it would follow that we also cannot be responsible for them. And if we are not responsible for our emotions, it seems unreasonable or even immoral to hold us accountable for them. Hence, we could never be justified to blame or resent someone for their emotions, no matter how misguided or even hazardous they are. Put more formally, this type of counterargument goes as follows:

1. It is only justified to blame people for something if they are responsible for it.
2. People are only responsible for something if they have control over it.
3. People don't have control over their emotions.
4. Therefore, people are not responsible for their emotions. (from 2 & 3)
5. Therefore, it is not justified to blame people for their emotions. (from 1 & 4)

Since this argument is valid and its conclusion contradicts my thesis, in order to defend my position, I have to propose some way of showing that

the argument is unsound. There seem to be several options if we want to deny the conclusion. I briefly discuss these options and point out some strengths and problems to motivate my own strategy.

First, we could deny Premise 3 that we don't control our emotions. There are attempts to account for control by way of reason responsiveness. An example of this approach is given by Hieronymi,[2] who argues that we do have a type of control over mental states, such as beliefs. However, such accounts often stretch the notion of control to accommodate aspects of our mental lives we don't experience as very controlled, like beliefs, desires and also emotions. In effect, we would have to give up a stronger notion of control in favour of one that can encompass both actions and mental states. However, a stronger notion of control remains useful, for example, in distinguishing between controlled and uncontrolled actions. Hence, we might want to resist giving up such a useful notion simply to bring emotions and beliefs into the scope of responsibility. I think it's a valuable thing to be able to distinguish, for example, uncontrolled from controlled bodily movements, not only for the sake of moral responsibility theories but also for action theory. Widening the concept of control and including weaker forms to encompass mental states like beliefs and emotions might obliterate this distinction.

Second, we could deny Premise 2 that control is required for responsibility. This is the approach that Smith takes, by keeping a more traditional account of control but denying that it is required to ascribe responsibility to a person.[3] However, this might lead to troubling implications for other forms of accountability. For example, one common justification for punishment is that it poses a deterrent or dis-incentive. But if we cannot control whether to risk punishment, such a justification seems to fail. If we want to use the notion of responsibility in a justification of punishment, we should not eliminate the control requirement. It would be unfair to punish people for things that are out of their control simply because our notion of responsibility doesn't require control.[4] I don't think we should accept a general notion of responsibility that has such an upshot – especially if we use it in the justification of such things as punishment.

Third, we could deny Premise 1, that to blame someone for something, they need to be responsible. While this option, at first glance, seems the most ridiculous, I think it offers a good start for a further discussion of what it means to hold and be held accountable. My focus here is not on finding a general notion of responsibility, but only on whether we can ever be justified in blaming others for an unfair emotion. Since the discussion of whether responsibility has a control requirement seems to lead to the difficulties mentioned earlier, it does not seem unreasonable to simply ignore it and focus on accountability. My proposal, therefore, is to cut

out the question of responsibility and directly ask: Is control required for accountability?

4.2 Disentangling the Control Requirement

What does it mean to hold someone accountable? There seem to be different ways to do so. For one, holding someone accountable for damages could mean to require them to pay for the repair or compensation. This seems to be an instance of a more general notion of holding accountable, whereby we make people take on the negative consequences of their actions. However, there are many cases where damages cannot be repaired or simply making people repay seems not enough. A common practice of trying to dissuade people from bad conduct is to put sanctions on it. For example, if we don't want people parking their cars in a specific place, we might put a fine on doing so. If someone violates the given norm, we hold them accountable by way of those sanctions. These are not in place to make the perpetrator bear the negative consequences of their own behaviour directly, but to put a high cost on a type of conduct to dis-incentivize it.

A third type of holding others accountable is reactive attitudes, such as anger, resentment and indignation. In order to resent someone for what they did, we don't need to enact any sanctions or act in any specific way. This type of accountability consists of a change of attitude on the part of the person holding accountable. Other forms of accountability might follow such a change of attitude. Authors often connect anger and resentment with a desire to punish or the opinion that punishment would be appropriate. But resentment neither requires punishing someone, nor does it constitute a form of punishment by itself.[5]

There are more accounts of blame, some of which don't easily fall into one of these categories. An example of this is the conative account by Scanlon[6] which shares some similarities with the reactive attitudes account but denies the necessity for these attitudes to be emotions. He instead views blame as a modification of one's relationship with the offender, and that this can be done purely by modifying one's desires or intentions, without a necessary emotional component. However, I limit my discussion here to these three types of accountability, since they provide three cornerstones for the space of common accounts of accountability.

Do all these ways in which we can hold someone accountable have the same responsibility conditions? Is there one relevant sense of moral responsibility that is required to justify sanctions, blame, and reactive attitudes like resentment, or do these accountability practices have different responsibility requirements, which again have different control requirements? With all these restrictions on the adequacy of a responsibility conception,

trying to find one single conception of responsibility that satisfies all types of accountability while remaining substantive and coherent seems to be a futile endeavour. Different schools of thought on moral responsibility have formulated appropriateness conditions for accountability. For a long time, philosophers have been arguing that some form of free will or voluntary control is necessary for moral responsibility; more recently, multi-level approaches such as that of Frankfurt[7] see intentions and second-order volition as relevant; and the RA theory views the agent's quality of will as the deciding factor.[8] Since so many different philosophical discussions draw on a notion of responsibility, I doubt that there is a single concept that can satisfy all, as illustrated in Figure 4.1.

For example, it might be that control is required for a specific concept of responsibility that is itself a requirement for one type of accountability practice. But at the same time, there might be another account of responsibility that is required by another type of accountability practice, but that does not itself require control. In such a case, it would be wrong to deny either of the two conditions. Responsibility as a concept is too dependent on the context of the discussion to deny either one in general. Hence, we would need to disambiguate different notions of responsibility. But if we simply disambiguate responsibility by the type of accountability practice we have in mind, responsibility simply becomes an intermediate step that does not add any independent constraint to the argument. Responsibility, in the sense required for punishment then just means a specific set of conditions for justifying punishment; and the responsibility required for resentment just means a set of conditions required to justify resentment.

As much as these theories of accountability differ regarding what conditions of justification they rely on, they also differ in what they seek to justify. RA theory sees ill will not as the basis to justify punishing someone, but only as a basis that justifies anger or resentment; and free will theorists don't usually consider the appropriateness conditions for resentment to be all that important. So it is unclear why we should require ill will on the

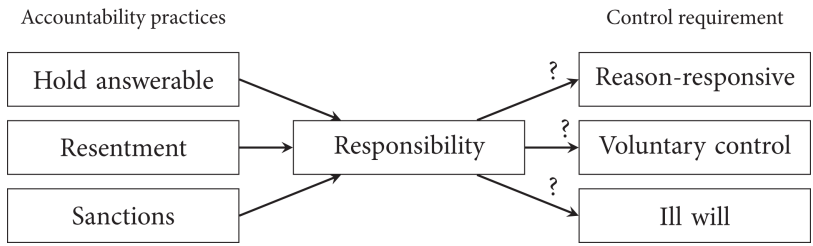


Figure 4.1 Responsibility as a singular concept with a unified control requirement.

side of the perpetrator to sanction a certain behaviour or free will to resent someone.

My proposal is to deny that we need a common notion of responsibility to all forms of accountability. We should rather disambiguate the control requirements based on what form of accountability we are discussing, as illustrated in Figure 4.2. As a rough proposal, we might want some form of voluntary control in order for sanctions to be appropriate. Sanctions seem to work by dis-incentivizing certain actions. Hence, in order to work, agents need to be able to weigh their options and decide against the sanctioned course of action. On the other hand, we don't need voluntary control to be justified in resenting someone. If resentment is a fitting response to someone else's lack of due regard or vicious attitude towards you, there might be no need for any strong notion of control. I will elaborate on this idea in Section 6.1.1.

If we get rid of the concept of responsibility for our argument, we need to collapse the first two premises and reformulate the argument in the following way:

1. It is only justified to blame people for something if they have the relevant type of control over it.
2. People don't have the relevant type of control over their emotions.
3. Therefore, it is not justified to blame people for their emotions. (from 1 & 2)

Whether this argument is sound depends greatly on what we think blame is, which type of accountability practice best captures the relevant notion of blame, and what type of control is required for that practice. In the following, I argue that the argument is plausible for some types of accountability practices such as sanctioning and punishment, but not for others, such as holding someone answerable or responding with reactive attitudes. I agree that to sanction or punish someone for something, some form of

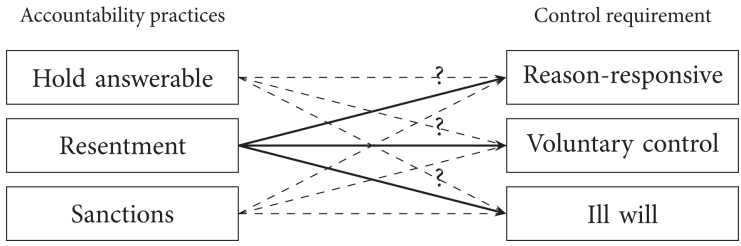


Figure 4.2 Control requirements disambiguated by type of accountability practice.

substantial control condition needs to be established. However, when it comes to reactive attitudes such as anger and resentment, we should reject Premise 1. It is not plausible that we need a strong type of control over our emotions in order for someone to appropriately resent us for them – but only a weaker requirement, such as reasons-responsiveness. In this chapter, I make the negative case why we should not assume a strong control condition for the justification of reactive attitudes, even though such a condition would be required to justify other forms of accountability. In the next two chapters, I make a positive case for what instead are sufficient conditions for appropriate resentment for emotions, without a strong control condition.

In Section 4.3, I elaborate on the possibility of denying Premise 2. I explore the idea that we are never in control over bringing about our emotions and identify the kinds of control we plausibly do have over our emotions. I argue that the available options pose a dilemma: either we have to broaden the notion of control and thereby lose some useful distinctions, or we can only rely on indirect forms of control over our emotions that are not sufficient to account for the accountability for unfair emotions themselves.

In Section 4.4, I disambiguate different forms of holding people accountable. My aim is to establish what kind of accountability practice amounts to blame. Specifically, I want to establish whether there is a conception of blame that can apply to emotions and which is a legitimate way of holding someone accountable for a moral wrong. To that end, I also examine what types of control are necessary for the different types of accountability practice.

4.3 Controlling Emotions

What kind of control do we plausibly have over our emotions? I first want to examine three types of control that we are likely to have over emotions. I group these types of control into three categories, direct voluntary control, indirect voluntary control, and rational control. Both direct and indirect *voluntary control* are types of control that we can consciously and deliberately exert over our actions and mental states. In the case of emotions, direct voluntary control is the least well-supported and most controversial. But there are nonetheless some considerations that speak in favour of people having some limited degree of direct voluntary control over their emotional episodes. Indirect voluntary control includes most types of conditioning, habituation, or therapy, that aim at influencing our emotional episodes long-term. Both of these are among the more classical interpretations of control and are often invoked by authors as requirements for responsibility or various forms of moral accountability.

The third category, that of *rational control*, differs from voluntary types of control in that it cannot be exerted deliberately or *at will* by a subject. Nonetheless, there are several accounts of control proposed by authors in the responsibility literature that fall into this category. These accounts can still be considered accounts of a type of control because they describe mental processes that are guided by the attitudes, capacities, and abilities of the subjects, similar to voluntary types of control. There are certainly arguments against calling this type of mental process a type of 'control'. However, I am less concerned with the terminological question here, but rather with the question whether such a type of rational control can sufficiently fulfil the *control condition*[9] for justified blame.

4.3.1 Direct Voluntary Control

Direct voluntary control over a mental state would mean that a subject can make a decision to create, get rid of, or change that state. Two aspects distinguish this type of control from the others discussed here. First, the process can be a conscious and intentional one. When a subject has this kind of control over a mental state, they can make a conscious choice or decision to change the mental state – although it does not always have to be the case. I am not going too far into the question of what the exact conditions for a conscious choice or decision are. But I think we commonly have a clear notion of when we have it and when we don't.

The type of control at issue here is the one we have over actions. For example, you exercise this type of control when you can raise or lower your arm *at will*. You can at a moment's notice decide to raise your arm and, barring any physical or medical restrictions, move it. When applying this picture to emotions, direct voluntary control would entail that all it takes to change your emotional state would be to have such a moment of choice or decision and then be able to shift your emotional state accordingly. This seems, at first glance, not possible for typical humans.

There are, however, several psychological techniques to regulate emotional responses at the moment they occur. For example, we can suppress an emotion to a certain degree, thereby diminishing its force or influence over our immediate mental state. We can also shift our focus away from the features that elicit the emotion, averting our eyes or thinking of something soothing or calming. This might similarly decrease the emotion's intensity. Alternatively, we can go through a process called *re-appraisal* to shift our perception and interpretation of the situation in such a way that the emotion seems no longer a warranted response.[10] This can also help decrease an existing emotional episode or even elicit a new one.

But do these techniques count as direct voluntary control of our emotions? It seems that there is indeed a moment of decision or choice in the

application of these techniques. Techniques like suppression or re-appraisal can be consciously deployed by the subject. We can, when we feel ourselves getting angry, decide to suppress the upcoming anger or give into it and let it grow in intensity. Alternatively, we can decide to avert our eyes in a scary film scene or think of something joyful when overcome with anxiety or sadness. But does this really constitute a direct voluntary control over the emotions themselves?

At closer inspection, these techniques seem somewhat limited to affecting the strength or duration of existing emotional episodes, and less effective in completely erasing an emotion or generating a completely new emotional state. Hence, they fall somewhat short of the *raising–your–arm*-part of the above picture. Averting your eyes or thinking of something nice, for example, does not change your existing emotional stance towards the object of your fear or sadness. If you look away, you might still fear the scary movie monster, but you simply happen to focus your attention on a different object, thereby replacing one dominant emotional episode with another one, which is directed towards a different thing.

A similar worry can be raised with the re-appraisal technique. What you are engaging in when re-appraising your situation is not changing your emotional stance at will, but looking for new information about, or a different plausible interpretation of your situation. If you fail to find any, you also cannot effectively change your emotion. However, even if you do find that the situation can be viewed in a different light, you cannot change your emotional stance to whatever you desire, but only to the type of emotion that is elicited by this new appraisal. For example, if you felt fear at the prospect of failing your education and then through a process of re-appraisal found that the system is rigged against you, you could not simply choose to feel relieved but would probably switch from fear of failing to anger over the unfairness.

Can these types of techniques match the direct voluntary control people can exert over their bodily movements? Suppression and attention shifting don't seem to do so. The most direct effect they can have on a subject's emotional episode is by up- or down-regulation of the emotional intensity. Also, shifting attention away from the object of your emotion does little to change that emotional episode, and more to shift from one episode into a different one with respect to a different object. Hence, the little amount of voluntary control people seem to have is over the intensity or duration of their emotional experience, and not over the type of emotion they experience. Additionally, the type of voluntary control we can exert is somewhat indirect, since, while it is possible to suppress an emotion or down-regulate it, it does not seem that we can freely choose to feel any given emotion about any given object.

4.3.2 Indirect Voluntary Control

Indirect voluntary control over mental states would entail that while we might not be able to change a state at will, we can take measures or engage in long-term practices that influence future instances of these mental states. In effect, indirect control is not a form of control over a specific instance of a state but much more over the likelihood of a similar kind of state occurring. We have several ways of indirectly affecting the emotions we are likely to feel, including therapy, habituation, avoiding trigger situations, training what we focus on, and more. This already establishes a certain type of indirect control over our emotions. There seem to be two types of indirect control that can be distinguished, indirect control over *specific* likely instances of an emotion – for example, by avoiding foreseeable trigger conditions – and indirect control over the *general* occurrence of a type of emotion – by changing emotional dispositions themselves.

In the first case, we might reasonably expect to feel a certain emotion in a specific future instance and take steps to prevent it, or conversely to make a future emotion more likely. For example, in a situation where you foreseeably feel xenophobic fear, you might not actually be in control over feeling the fear at that moment, but rather only over such factors as to avoid such situations or going to therapy. However, these failings don't seem to address the core of what is wrong with the unfair fear. In the case of unfair fear, what makes up the unfairness is the action tendency to avoid or not associate with the target of my fear for either mistaking the situation, in the case of a misplaced emotion, or due to vicious underlying evaluative attitudes, in the case of a misguided emotion. Xenophobic fear seems to fall more clearly into the second group, being based on prejudice or harmful stereotypes about people from different cultural backgrounds. Simply avoiding being confronted with people of other backgrounds does not seem to remedy that. Hence, what you can indirectly control is not actually your tendency to feel such an emotion when confronted with its trigger, but merely the likelihood of such a confrontation occurring.

In the second case, we might not have any specific future instance in mind, but strive to change our emotional dispositions in some way. For example, you might join a self-help group or start a meditation practice to reduce your tendency to get angry so easily. Or you might go to therapy to alleviate your fear of flying or spiders. Moreover, D'Arms argues that an additional form of influence we can have over our emotional reactions is by long-term training or exercise of our emotional sensibilities.[11] This can also amount to training such sensibilities like gaining a better sense for the quality of wines, a more refined sense of humour, or also a keener sense for injustices. This is a type of control that gets closer to the heart of the issue,

namely, changing the subject's underlying sensibility. However, it does not establish control over the immediate emotional episode either.

4.3.3 Rational Control

Rational control consists in an influence over a mental state through a mental process that does not have to be voluntary or consciously guided – meaning that there is no conscious choice or decision involved. This category includes accounts of what type of influence or control is required for responsibility such as *reason responsiveness*,[12] *evaluative control*,[13] and *rational relation*[14] or *judgement sensitivity*.[15] What these accounts have in common is that they don't focus on an instance of choice or decision between courses of action, but rather on the relation between the object of responsibility and the reasons an agent has for it.

Hieronymi distinguishes two types of control we have over our mental states, such as beliefs or emotions. While we do not have direct voluntary control over many of our mental attitudes, we can exert a type of indirect *managerial control* and a more direct *evaluative control*. Managerial control, which I would categorize as an account of indirect voluntary control, consists in the ability to intent to change or manipulate our attitudes and then take steps to implement these changes. Hieronymi argues that we can often take steps to indirectly influence our beliefs, for example, by deciding to investigate the matter. However, this ability is limited by whether we come to find good reasons for the desired beliefs.[16]

Evaluative control, on the other hand, is the ability to form an attitude on the basis of the right kind of reasons for that attitude. This applies to beliefs, but also to intentions and potentially other mental states. The idea is that there is a class of reasons – *intrinsic* or the *right kind* of reasons – that are valid reason for a belief. All others are *extrinsic* or the *wrong kind* of reasons and are not actually reasons to believe at all. In the case of beliefs, this means that we can only evaluatively control our adoption or rejection a belief based on evidence for or against the content of that belief. We cannot evaluatively control beliefs for other, so-called *extrinsic* reasons. While evidence gives you an intrinsic reason to believe something, someone paying you a lot of money to believe something does merely give you a reason to want to believe. Unfortunately for your finances, beliefs are the type of mental state that you can only form based on the right kind of reason – evidence – and not wrong kinds such as bribes or convenience. The same is true for intentions. We can only evaluatively control an intention to do something for the right kinds of reasons, and therefore, similar limitations count for beliefs as they also do for intentions.[17]

If we have the same type of evaluative control over our emotions, then it needs to be the case that there are right and wrong kinds of reasons for

emotions, something that authors writing on the fittingness of emotions have widely been claiming.[18] The idea is that there are fittingness conditions that provide the right kinds of reasons for a given type of emotion, just like evidence provides the right kind of reason for a belief, or certain practical reasons provide the right kind of reasons for intentions. These fittingness conditions vary depending on the type of emotion in question. Fear, for example, is fitting in the face of danger, sadness if something is a loss or anger at an offence or injustice. Evaluating a situation as matching these conditions by way of experiencing the corresponding emotion would then count as exerting evaluative control over your emotions – meaning responding to the right kind of reasons with the appropriate mental state.

Smith similarly describes a type of influence we can exert over mental states, like beliefs or emotions, which she calls *rational relation*[19] or *judgement sensitivity*.[20] She contrasts this type of rational influence with voluntary control to make the point that we don't need voluntariness to be answerable for mental states, such as beliefs or emotions. This type of influence relies on a close rational relation between the attitude, like an emotion, and underlying evaluative judgements.[21] For example, when you feel fear of a spider, this emotion is rationally connected to your underlying judgement that the spider poses a threat to something that you value, such as your health. This connection is such that it can sometimes fail or misfire, but that if you hold the underlying evaluative judgement, then you should rationally also feel the emotion. The fact that this connection can sometimes misfire does not mean that there is no rational connection. Rather, that the connection often does hold and that we clearly recognize it as a rational failing when it misfires shows that there is such a rational entailment.

4.3.4 Rational Accessibility of Emotions

Whether our emotions are open to rational control or not is not a clearly settled question. Not all theories of emotion allow for the same responsiveness to reasons in the generation of emotions. The two main theories that might face problems with the idea that emotions are open to rational control are somatic and perceptual theories of emotion. Both draw analogies between emotions and sensory experiences – portraying emotions either as the sensory experience of the state of one's own body in somatic theory,[22] or as analogous to the perception of the object of the emotion.[23] But neither of these types of sense perception seems very open to influence by reasons.

For one, if you feel pain, you cannot stop feeling it even if you have good evidence that nothing in your body is damaged. Or if you feel cold, having good reasons to think that the room temperature is reasonably warm does not reliably change the fact that you feel cold. Similarly, even if you

have good reasons to believe that you are having an optical illusion, that does not reliably change the illusion itself. For example, if you perceive a stick in a glass of water bending at the surface, knowing that the stick is straight and that the illusion is created by the refraction indices of air and water does not change that you see it as bent. Therefore, if emotions are analogous to sense perceptions of either of these types, they might also be inaccessible to rational considerations.

The phenomenon of recalcitrant emotions speaks in favour of this inaccessibility. A recalcitrant emotion is an emotion you feel, for example fear of a spider, despite believing that there is no cause for it, like knowing that the spider is harmless. If emotions were accessible to reasons in the same way as beliefs are, then such a recalcitrant emotion would be clearly irrational. There is a larger debate on whether recalcitrant emotions are in fact irrational or what kind of rational fault is involved in them. Positions range from denying their irrationality and seeing them as analogous to optical illusions[24] to ascribing them irrationality analogous to contradictory beliefs.[25, 26]

However, while it makes little sense to ask for reasons *why* someone perceives a tree, asking for reasons *why* someone feels angry or afraid seems perfectly intelligible and common-place.[27] Even if we accept that emotions are in some aspects meaningfully analogue to perceptions, and thereby largely unaffected by reasoning processes, they might still in some ways differ from perception – for example in the possibility of posing such *why-questions*. Emotions seem to depend to a relevant degree on the subject's other mental states. For one, Döring argues for a perceptual theory of emotions in which emotion nonetheless represent their objects in light of the subject's concerns.[28] For example, while most humans might fear something that threatens their health and safety, not everyone fears that a treasured national tennis star might lose the next Grand Slam tournament. To fear such a thing, you first need to care about that person's sports career, or at least about what happens in the world of tennis. Similarly, Callard argues that many of our emotions reflect the things we value, giving examples like the sadness at the prospect of emigration being a sign of love of one's homeland, or fear of a low grade reflecting one's commitment to academic achievement.[29] Hence, at least some of your emotions about specific things reflect what things you care about on a more general level.

The connection between underlying attitudes and specific emotions, I argue, is just as much a type of judgement sensitivity as the ones discussed by Hieronymi and Smith. It is through a rational capacity that we can recognize the relevance of an object for our underlying cares and values. The specific way in which it is relevant to our cares and values is what determines that type of emotion we consequently feel. For example, if the object threatens what we hold dear, we would respond with fear; if it disregards or offends what we hold sacred, we respond with anger; and if it aids in

bringing about what we hope for, we might feel excitement towards it. On the flip side, we can be mistaken about the relevance of an object for our underlying values and mistake a harmless spider as a threat to our health or a mere statement of fact as an offence to our cherished beliefs. In such cases, the resulting emotions would still reflect the underlying values, but lack the rational support that would have been given if the objects were actually relevant in the perceived way. This connection can hold even for theories of emotions that compare them closely to perceptions, as can be seen in the case of authors such as Döring.[30]

4.4 Ways of Holding Accountable

To decide whether it is ever justified to blame someone for their emotions, we need to clarify the relevant notion of blame. What we are looking for is a notion of blame that can under certain conditions be an appropriate response to a moral wronging. We then need to establish what kind of control requirement this notion of blame entails and whether we in general have the relevant type of control over our emotions. Broadly, blame is a negative reaction someone else has to a moral wrong committed by the blamed person. It is a type of holding that person morally accountable for the wrong. The following three features are useful to outline the general concept of blame.

First, blame is an action or attitude that one person performs or holds towards another. While there are many differing accounts of blame, they all involve some type of mental state or overt action by someone. To blame someone might mean to hold a judgement that they are causally responsible, to feel an emotion like anger or resentment towards them, or even to perform a speech act such as accusing them. This distinguishes blame from mere blameworthiness. While someone being blameworthy does not imply that there is anyone who actually blames them, blame itself is not hypothetical in this way. For example, if you were the last person in the world because you killed the second-to-last person in the world, there would be nobody left to blame you – with the exception maybe of yourself – but you would still be blameworthy.

Second, blame is negative. While there are positive forms of holding people accountable, by praising them or admiring them, blame is a negative type of holding someone accountable. This can mean two things, both of which apply to blame. For one, blame is a type of holding people accountable for negative things, like wrongful actions – or potentially wrongful attitudes. For another, blame is somehow negative in itself. This can mean that blame is either an attitude or type of behaviour that is uncomfortable or undesirable for the person being blamed or that it is even uncomfortable to be blamer.

Third, blame implies a normative upshot for the blamed person. Meaning, if blame is appropriate then it follows that the blamed has reason to answer for themselves – to apologize, to make reparations, or to excuse or justify their behaviour to the blamer. This can be understood in at least two ways. For one, if you are being blamed for something you have actually done wrong, you might have good normative reasons to answer for yourself. For example, we might have a moral duty to apologize to someone we have wronged. For another, there might be more or less clear social expectations and conventions regarding how you should behave when being blamed. Such conventions might either carry little moral weight by themselves or they might simply capture the moral implications of wrongdoing or specify the conventional way of how an apology or excuse is made in the current society. For example, you might be required to pay the wronged party or offer them a gift, or it might suffice to express your sincere remorse.

These three features not only are marks of blame but also capture the kind of complaint that the target of an unfair emotion can sometimes have against the subject. The judgement that an emotion is unfair characteristically involves or at least supports such a negative attitude towards the subject and the expectations that they answer for their unfair emotion. This is, of course, not the case if it turns out that the subject is not at fault for the emotion – for example, in the case of a misplaced emotion that is based on a faultless mistake.

Since there are numerous accounts of blame[31] and even more of moral accountability[32] in general, I have to limit the discussion here to a manageable selection. I focus on four general notions of blame and accountability, which go by several different names, but which I will call *attribution*, *answerability*, *social sanction*, and *reactive attitudes*. These four notions cover a large territory, from very minimal accounts of blame, like attribution or answerability, on the one side, to much more engaged or impactful accounts of blame, like social sanction, on the other side. While there are other accounts that could fall outside these notions, I find it useful to examine the extreme positions and to contrast them with a popular intermediate one – here the reactive attitudes theory of moral accountability.[33]

In Section 4.4.1, I argue that while it is highly plausible that we can attribute emotions to a subject, and do not require any direct voluntary control over our emotions, attribution does not sufficiently constitute an instance of blame. Therefore attributability cannot explain the special type of complaint that the target of an unfair emotion can sometimes make against the subject.

In Section 4.4.2, I discuss a slightly more demanding notion of blame, *answerability*, which is nonetheless on the low end of both control conditions and impact on the accountable party. I follow Smith[34] in her argument

that the type of control required for answerability is rational control. Furthermore, it is highly plausible that we, in principle, have such a type of rational control over our emotions and therefore are sometimes answerable for them.

In Section 4.4.3, I turn to a much more forceful and impactful notion of blame, in the form of social sanctions. I argue that while this is not an unreasonable notion of moral accountability, it does not apply to emotions. The type of control required to justify social sanctions has to be in the range of direct voluntary control, something we largely lack with respect to our emotions. However, while sanctions can be a form of negative moral accountability practice, it is not the only one.

In Section 4.4.4, I argue that responding with reactive attitudes, and more specifically resentment, is both a viable account of blame and one that can apply to emotions. I elaborate on the details of how and when it is appropriate to resent someone for feeling an unfair emotion in Chapters 5 and 6.

4.4.1 Attribution

The idea of attributability is that some actions, attitudes, or traits are connected to a person in a more relevant way than others. For example, while your height is a feature of you, it is in most cases not a result of your decisions, deliberations, or reasoning. In contrast, when you choose to get a tattoo as a form of self-expression, that is very much attributable to you because it is an expression of your desires, values, and reasoning process.

In the philosophical literature, attributability is often used to distinguish autonomous actions from actions under coercion or manipulation by others. It is a useful concept to express the idea that some actions are more of the agent's own doing or choosing, while others are due to external constraints. However, is attribution the type of attitude that can count as a viable conception of blame? Watson distinguishes attributability in this sense from accountability as two distinct, but independently important notions of responsibility.[35] Attribution is thereby a distinct phenomenon from holding someone morally accountable and also the notion of blame. The fact that something is attributable to a person does not, by itself, answer the question of normative upshot, that is, if there are any claims or obligations for that person that are raised as an implication of the attribution.

It is true that attributing an action or attitude to another person can be seen as a relatively weak type of accountability practice. It does fit the first feature of blame, namely, that it is an attitude that one person can hold towards another. I can, for example, attribute your actions to you and see them as a proper consequence of your decisions, deliberations or reasoning, free of coercion or manipulation. And there have been accounts of

blame proposed that are closely related to the notion of attribution. Smart, for example, understands blame as grading and an ascription of responsibility.[36] In this sense, it is merely a judgement or attitude of attribution that involves a morally negative grading of the blamed person.

However, several authors highlight the depth and negative aspects of blame. These authors criticize the pure attribution of a moral wrong as too weak to explain the force and depth of blame.[37] Hence, attribution fails on the second feature of blame, since it is not necessarily a negative attitude. It could be argued that only some instances of attribution would count as viable interpretations of blame, namely, that only attributions of actions or attitudes judged as morally objectionable to a person would count as a type of blame. However, even in these instances, the attribution is only negative in the first sense described earlier – that what is attributed to the subject is evaluated negatively. The attitude of attribution itself is not necessarily negative in the sense of feeling bad or uncomfortable to the blamer or the blamed.

On the third feature, normative upshots such as a duty to apologize or make amends, attribution by itself seems to be completely neutral. It is neither obviously the case that you ought to apologize for something that is attributable to you, nor that you don't need to. For these reasons, I follow Watson[38] in not counting *mere* attribution as a form of holding accountable, although it might be a necessary condition for some other forms of accountability practice. In any case, attributing an emotion to a subject is not strong enough to count as blaming them, nor does it meaningfully illuminate any special standing to blame.

4.4.2 Holding Answerable

The next way of holding accountable that I want to discuss is the practice of holding people answerable. When we hold someone answerable for something, we hold them to the expectation that they can justify the thing we hold them answerable for, that is to give us reasons for it. For example, imagine you are showing a flower you picked to a friend and they immediately slap it out of your hand. When you in response ask them why they did that and insist on an answer, you're holding your friend answerable for their action. This does not have to include being angry about what they did or imposing any sort of consequences or sanctions on them, but simply insisting that they give a reason for their behaviour. Holding someone answerable like this can be seen as a way of holding them accountable because it is not simply a judgement we make privately but a social practice; by asking for justification or otherwise making it clear that we expect a justification from someone, we engage in an interaction with that person that is about the action or attitude at issue.

The primary proponent of an answerability account that allows for people being answerable for their emotions is Smith.[39] According to Smith, mental states are attributable to a person if and only if the person is *answerable* for them. This account allows for more than merely forming an attitude of attribution, as discussed in the last section, but also enables us to ask the answerable person for reasons for their attitudes or behaviour. Being answerable for an attitude or mental state means that we can in principle give reasons for them, and justify them to others. The condition for this type of answerability to apply to emotions is therefore that our emotions be responsive to reasons, and thereby products of a rational mental process. That is to say, we can ask questions such as 'why did you get angry?', asking for reasons for the anger, rather than simply a causal explanation. Being in a position to answer this kind of questioning then means that the emotion is also properly attributable to you.

This type of answerability, unsurprisingly, requires that we can exert rational control over the mental state in question – for emotions to be reason-responsive mental states. This is because, to be able to give reasons, we need to have reasons for the state, and this requires the capacity to form an emotion in light of our other mental states, such as judgements, beliefs, or underlying values. This does not require the ability to exert voluntary control over our emotions, in order to be answerable for them. Requiring voluntary control to simply be answerable, meaning able to provide the reasons for which we have arrived at a mental state, would be superfluous. The only type of reason that would be added by the ability to voluntarily control something are reasons that don't actually speak for the emotion itself, but would be *extrinsic* reasons.[40] For example, if I asked you why you believed that the earth is round, the relevant reasons would be those that constitute evidence for your belief – *intrinsic* reasons. Being able to provide such intrinsic reasons is enough for being answerable. Your lack of voluntary control over this belief would mean that you could not change your belief for an extrinsic reason like the one that you would like to live on a flat planet. And if I asked you for a reason for your belief, and you said you simply liked it more, that would not count as being able to provide an actual reason for the belief. Hence, the ability to give reasons for why we feel an emotion is not hampered by the inability to form it on the basis of the wrong kind of reasons in any way.

Since answerability is merely the condition under which we can intelligibly ask for the reasons an agent has for the relevant action, state or attitude, and not the act or attitude of holding them accountable, it does not count as blame. The relevant object of investigation therefore is not *answerability* but an act or attitude of *holding someone answerable*. The difference again being that someone could be answerable without anyone holding them answerable. What does it involve to hold someone answerable for their

attitudes or emotions? Either overtly demanding explanation or justification or having an expectation that they provide such. While this seems a reasonable interpretation of a way of holding someone accountable, it would be somewhat of a stretch to see it as an account of blame. Smith herself does not see answerability in this way either: she defends a different account of blame in which blame is primarily identified by its function as a type of protest.[41] Although Smith views answerability as the central notion of moral responsibility, she does not go so far as to say that answerability for mental attitudes by itself settles the question whether we are ever to blame for them.[42] However, holding someone answerable either by making the explicit demand for justification or by having an expectation is a viable way of holding someone accountable for their attitudes – one that readily can be defended as an appropriate way of reacting to someone's unfair emotions.

Consider, for example, your flatmate being angry with you. In this situation, demanding or expecting them to give a reason for their anger is a way to acknowledging that they carry a certain responsibility for their anger. The fact that you hold them answerable in this way shows that you don't see their emotion as something that merely happens to them but rather as an expression of their rational agency. Of course, they might have an irrational fit of anger that they cannot answer for. But that possibility does not preclude that in many cases they can give reasons and do accept their anger as their own.

Is this the relevant sense in which we can sometimes blame people for their emotions? I think not. Answerability only establishes that we are liable to having to justify our actions or attitudes, but does not yet establish either that the answerable person is at fault, nor the normative upshot that the answerable person needs to apologize or otherwise make amends. What seems to be missing is an element of condemnation or protest characteristic to blame – its negativity. You can hold someone answerable without a negative judgement, and while still being open to the possibility that they have done nothing wrong. You cannot, however, blame someone without such a negative judgement.

Holding someone answerable does however establish the sort of interpersonal attitude or relation that we are looking for. When we are answerable for something, we have an obligation to justify it to others. Answerability can even be directed at a specific person. In the flatmate example, if your flatmate owes anyone an explanation it is you. If some third person demanded that your flatmate answer for why they are angry, it would not seem out of place to answer with a curt 'None of your business!'

Feeling an emotion like anger towards someone that potentially poses a moral hazard to the target does seem to establish this relation of answerability between the subject and the target. But while this way of holding

the subject morally answerable for their emotions seems reasonable, it is neither limited to unfair emotions, nor does it have the negativity or the normative upshot characteristic of blame.

4.4.3 Social Sanctions

Some authors have interpreted our practices of holding each other morally accountable primarily by way of social sanctions.[43] A sanction is the imposition of a negative consequence on the agent in response to them performing an undesired action. Sanctions work best when they are known and thereby pose a deterrent for people who might have engaged in such undesirable behaviour. Therefore, while the sanction itself happens after the fact, and in response to the wrongdoing, the threat of sanction already impacts informed agents' decisions before they act.

Is control required for a sanction view of blame? It seems plausible that it is. The point of sanctions is either to make an offender change their current behaviour or to make them not repeat their bad behaviour in the future. For sanctions to work, they need to be able to affect a person's deliberation about what to do and then act based on those deliberations. To act in this deliberate way clearly requires some type of voluntary control. However, one conceptualizes voluntary control, acting out of the conclusion of one's deliberation is surely a paradigmatic case of it.

This has two problems when applied to emotions. First, it assumes a type of voluntary control that we simply don't have over our emotions. As we have seen, while we might have some indirect form of voluntary control over our emotions, by avoiding trigger situations or through habituation, we mostly can't feel at will. In effect, while sanctions can incentivize people to try to habituate or otherwise indirectly influence their emotions, it cannot incentivize people to feel or not to feel directly. As a result, there are always possible situations in which someone did not have the chance or the time to deploy these indirect measures and hence comes to feel a sanctioned emotion. In such cases we would either have to sanction them nonetheless, without giving them a chance to avoid the sanction – which seems unjust – or admit that what we actually want to sanction is the failure to indirectly control the emotion if one has had such a chance. Therefore, sanctions either are unjust or don't actually apply to the emotions themselves.

Second, sanctions provide the wrong kinds of reasons to feel or not feel an emotion. Sanctioning a certain emotional response or type of emotion only gives us *extrinsic* reasons or reasons to not want to feel those emotions. These must be distinguished from *intrinsic* reasons that actually speak for or against the emotion itself.[44] For example, if you are afraid of heights but all your friends want to climb up a high tower. Say you would

be socially sanctioned in the form of exclusion and being laughed at if you didn't join. In this case, you can go with them despite your fear, or try to suppress the fear, or distract yourself from the deep drop behind you. However, the threat of exclusion cannot give you an intrinsic reason to stop fearing, since it in no way changes the fearsomeness of the height for you. In effect, you cannot avoid feeling the emotion for the reason that we would otherwise incur the sanction. Contrast this scenario with a sanction on action. By putting a cost on an action, a sanction provides the right kind of reason against performing that action, because actions, in contrast to emotions, can be performed due to the balance of benefits and costs they incur. Sanctions work by directly affecting the agent's intrinsic reasons for the action, by changing the overall benefits of performing an action to an overall cost. But they cannot affect an emotion in the same way, since the costs and benefits of having an emotion are not part of the intrinsic reasons for the emotion. Hence, sanctions cannot work on emotions in the same direct way as they do for actions.

Simply suppressing unfair emotions or avoiding trigger situations does not address the problem at the heart of the unfairness, namely, that, in the most pernicious cases, the emotion is based on a vicious evaluative attitude of the subject. Reasons based on possible sanction don't directly address that issue, they merely incentivize the subject to avoid certain outcomes like getting into a situation where they might feel the emotion, which can also be accomplished by simply avoiding such trigger situations. And even if sanction can incentivize going to therapy or otherwise indirectly influence your emotional dispositions, they would still only give you reasons for these emotion-related courses of action, not for or against the emotions themselves. Also, in situations where an unfair emotion is the result of a prior mistake in interpreting the situation, being incentivized to not feel certain emotions does not really help clear up the mistake.

Some authors have argued that we don't need voluntary control to be blamed for mental states and attitudes because blame itself is not a form of sanction or punishment.[45] Carlsson counters this line of reasoning by shifting the focus from blame to guilt.[46] Even if blame itself is not a form of sanction or punishment because the target of blame does not necessarily suffer from it, feeling guilt is necessarily painful. If this is the case, and blame aims at, or at least is satisfied by the target feeling guilty, then any successful instance of blaming necessarily involves suffering. This reintroduces a form of sanction and thereby the necessity of voluntary control to justify blaming someone.[47]

According to a sanction view, blame and praise primarily serve the function to encourage people to behave in morally valuable ways and dis-incentivizes bad behaviour. This works only if being blamed feels bad in itself or comes with a range of behavioural changes. For example, when

I blame a person, I might not help them out in little ways, avoid or even shun them, and withhold many of the benefits of being part of a social group. But if we follow this line of argument, then it would again seem implausible that we would ever be justified in sanctioning people for their emotions. So, if a sanction view turned out to give us the best theoretical conception of blame, it would follow that we are never justified in blaming people for their emotions. If sanctions don't provide the right kind of reasons for or against emotions, then we cannot justify sanctioning people for what they feel, even if those feelings are repugnant to us.

However, the sanction view might not give us the best theoretical account of blame after all. Wallace provides several points of criticism of the sanction view of blame.[48] Specifically, there are two points that are useful for our discussion of blame for unfair emotions. First, the sanction view cannot explain private blame. While it is reasonable to connect blame to sanctioning behaviour, to directly identify the two makes it impossible to blame someone without resorting to outward, sanctioning behaviour. However, we do commonly blame people without going so far as to sanction them. There seems to be an attitudinal dimension underlying our blaming practice.

Moreover, it seems that making certain actions more costly without such an attitude does not actually amount to a sanction. For example, if you were prone to fits of anger, the more such episodes you had in the presence of others, the more they might start avoiding you. This could be interpreted as a type of sanctioning behaviour, especially when their avoidance seems like shunning or social exclusion to you. But simply reacting with avoidance to uncomfortable situations does not amount to blaming you for them. What is lacking in such a picture is an attitude of condemnation that motivates the avoidance behaviour.

Second, sanctions, or at least the threat of sanctions, are future-directed. They aim at changing a person's behaviour by placing high costs on undesirable behaviour. This means that even if the sanction itself is in response to a wrong, and thereby a reaction to something in the past, the way it works is by deterrence. There might still be good reasons to implement a sanction after the wronging has already occurred, even though its main purpose – to deter the action – has therefore already failed. For one, going through with the sanction signals that these are not empty threats, ensuring that other, future instances of similar actions are being deterred. Other than this secondary effect, enacting a sanction after it has already failed to deter an action might give people a sense of the wrongdoer getting what they deserve.

Blame, in contrast, is backward-looking, meaning it focuses on the moral wrongness of past wrongs rather than on future improvement. If blame is to provide a way of identifying that someone was wronged by an

unfair emotion, then it also needs to primarily be about that specific unfair emotion – and not, for example, about similarly unfair emotions that potentially might occur in the future. If the purpose of blaming someone after the fact is to signal the seriousness of the threat of sanctions or satisfy some sense that the offender deserved retribution, then it is no longer actually about the wronged party. Blaming someone for a wronging means blaming them on behalf of the wronged party, not on behalf of potential future wronged persons, and neither merely for the sense of satisfaction that the offender got their due.

4.4.4 Resentment and Reactive Attitudes

Holding someone answerable and attribution are too forceless, social sanctions are too demanding, but resentment, as I will show, gets it just right. Resentment is itself an affective attitude – an emotion or sentiment – towards a person. As an affective attitude, resentment does not necessarily involve overt sanctioning or punishing actions, but it involves action tendencies and a motivational force that is lacking in mere attribution. In contrast to holding someone answerable, resentment has a clear negative valence and is not neutral with respect to judging the resented person, but rather involves an element of condemnation and protest.

If this is the case, and resentment is rather a form of protest than just another form of sanction or punishment, then it does not necessarily have the strong control conditions as the social sanctions account for blame. If we understand it as an attitude of protest, then it still requires some type of control – like rational control. Protesting someone's eye colour or some other aspect they have no rational control over might be possible, but misses the point of protest. Protest, following Hieronymi and Smith, is a way of drawing attention to and challenging unacceptable claims made by people – made either explicitly or implicitly by their behaviour.[49] This implies that to properly protest another person's actions or mental states, you need to be able to understand the claims implicit in them – understand the reasons they have for them.[50] This is only possible for actions or mental states for which the agent has such reasons. Hence, one can only protest things that are sensitive to reasons – that are open to rational control.

However, Smith does not go so far as to argue that rational control over mental states or attitudes is *sufficient* for blame and moral protest.[51] She merely argues that it is sufficient for us to be answerable for these mental states. As discussed in the last section, I agree with this assessment if we understood blame as a type of sanction – or even overt forms of punishment. For such accountability practices, those subjected to them need to be able to choose and make the decision to avoid the sanctions. However, resentment and reactive attitudes do not require this type of voluntary

control. Therefore, I want to go a step further than Smith and claim that rational control is sufficient for the control requirement of resentment and other reactive attitudes; and that it therefore applies to emotions as well as actions. I come back to this point in Chapter 5, Section 5.2.2 and following sections.

There are two aspects of reactive attitudes theories that work well with my notion of unfair emotions. First, RA theories often focus on *quality of will* as the underlying issue with blameworthy actions. This nicely corresponds to a vicious underlying evaluative attitude in the case of a misguided unfair emotion. Second, RA theories mostly see resentment or anger not as a tool for sanction or punishment, but as a shift in the relationship between blamer and blamed, which serves a communicative function, for example, to protest the blameworthy action. These two points together show how blame can give the subject of an unfair emotion the right kinds of reasons to rethink and change their emotional reactions. Blame as protest does not signal the subject 'you better stop feeling that emotion or else ...!', as a sanction might, but rather that there is something intolerable in how they view and how they relate to the people they have wronged.

In conclusion, if we want to find a theoretical account for our practice of holding people accountable for unfair emotions, then we need to accept something like the reactive attitudes account as a valid form of holding accountable. For one, there are several independent reasons that speak in favour of reactive attitudes account of blame – for example, its backwards-looking nature and the possibility of private blame. For another, in contrast to a social sanctions view of blame, an RA theory of blame can account for the phenomenon that we sometimes blame people for their unfair emotions. Given these two points, I continue for the next two chapters to elaborate on how reactive attitudes, and specifically resentment,[52] can be a fitting and morally appropriate way to respond to unfair emotions.

Notes

1 Fischer and Ravizza, *Responsibility and Control*; Frankfurt, "Alternate Possibilities and Moral Responsibility."
2 Hieronymi, "Responsibility for Believing."
3 Smith, "Responsibility for Attitudes," 265, calls the requirement for attributability "rational control" once, but otherwise refers to it as *rational relation*. Hence, it is unclear whether she would fully reject (2) or be much closer to Hieronymi's account.
4 We could also deny the control requirement for responsibility but reintroduce it as an additional requirement for punishment. In that case, the relevance of responsibility in the justification would be diminished as well, raising the question of why we need a responsibility requirement to justify punishment at all. This comes close to the strategy I ultimately propose.

5 This point has also been argued by Hieronymi, "The Force and Fairness of Blame"; Smith, "Moral Blame and Moral Protest"; Wolf, "Blame, Italian Style."
6 Scanlon, "Moral Dimensions"; Scanlon, "Giving Desert Its Due."
7 Frankfurt, "Alternate Possibilities and Moral Responsibility"; See also Fischer and Ravizza, *Responsibility and Control*.
8 See, e.g. Strawson, "Freedom and Resentment," 1962; McKenna, *Conversation and Responsibility*; Shoemaker, "McKenna's Quality of Will."
9 For ease of argument, I will continue to call it the control condition, even if it can be satisfied by something that is not strictly a type of control.
10 For an introduction to the psychological literature on emotion regulation, see Gross, *Handbook of Emotion Regulation*; Gross, "Emotion Regulation."
11 D'Arms, "Value and the Regulation of the Sentiments."
12 Fischer, *My Way*; Fischer and Ravizza, *Responsibility and Control*; Scanlon, "Moral Dimensions."
13 Hieronymi, "Controlling Attitudes."
14 Smith, "Responsibility for Attitudes."
15 Smith, "Attributability, Answerability, and Accountability."
16 Hieronymi, "Controlling Attitudes," 54–55.
17 Hieronymi, "Controlling Attitudes," 59.
18 There is however a debate on whether these fit-making conditions should be properly called reasons or not. Proponents of the affirmative position include de Sousa, *The Rationality of Emotion*; Nussbaum, "Emotions as Judgments of Value and Importance"; Silva, "The Rationality of Anger"; Na'aman, "The Rationality of Emotional Change"; For a prominent argument against fit-making properties being genuine reasons, see Maguire, "There Are No Reasons for Affective Attitudes."
19 Smith, "Responsibility for Attitudes."
20 Smith, "Attributability, Answerability, and Accountability"; Smith, "Responsibility for Attitudes," 265, also calls this type of influence *rational control*, however she mostly refrains from describing it as a type of control.
21 Smith, "Responsibility for Attitudes," 253.
22 See, e.g. Prinz, *Gut Reactions*.
23 See, e.g. Tappolet, *Emotions, Values, and Agency*.
24 Döring, "Why Recalcitrant Emotions Are Not Irrational."
25 Nussbaum, "Emotions as Judgments of Value and Importance."
26 See also Brady, "The Irrationality of Recalcitrant Emotions" for an intermediate position.
27 Deonna and Teroni, *The Emotions*, 69.
28 Döring, "Seeing What to Do," 385.
29 Callard, "The Reason to Be Angry Forever," 127 ff.
30 Döring, "Seeing What to Do."
31 See Tognazzini and Coates, "Blame," for an overview of the current debate.
32 See also Talbert, "Moral Responsibility."
33 Scanlon, "Moral Dimensions," chap. 4, also proposes an intermediate account of blame that involves a change in attitude and intentions towards the blamed person, however, he does not identify this change as a form of reactive attitude, or as necessarily involving any emotions.
34 Smith, "Responsibility for Attitudes."
35 Watson, "Two Faces of Responsibility."
36 Smart, "Free-Will, Praise and Blame," 305.

37 Wolf, "Blame, Italian Style"; Hieronymi, "The Force and Fairness of Blame"; Wallace, "Dispassionate Opprobrium."
38 Watson, "Two Faces of Responsibility."
39 Smith, "Responsibility for Attitudes"; Smith, "Attributability, Answerability, and Accountability."
40 See Section 4.3.3.
41 Smith, "Moral Blame and Moral Protest."
42 Smith, "Responsibility for Attitudes," 266.
43 See, e.g. Dennett, *Elbow Room*.
44 The right versus wrong kind of reasons distinction is often also made for beliefs and argued for by authors like D'Arms and Jacobson, "Wrong Kinds of Reason and the Opacity of Normative Force"; Hieronymi, "The Wrong Kind of Reason"; Gertken and Kiesewetter, "The Right and the Wrong Kind of Reasons."
45 See, e.g. Hieronymi, "The Force and Fairness of Blame"; Graham, "A Sketch of a Theory of Moral Blameworthiness."
46 Carlsson, "Blameworthiness as Deserved Guilt."
47 Carlsson does not further specify why pain or punishment requires voluntary control, but merely assumes it as an intuitively plausible premise.
48 Wallace, *Responsibility and the Moral Sentiments*, 55 ff.; Wallace calls this view of moral blame the "economy of threats" view, a term coined by Hart, *Punishment and Responsibility*.
49 Smith, "Moral Blame and Moral Protest," 42–44; Hieronymi, "Articulating an Uncompromising Forgiveness," 546 ff.
50 Hieronymi, "Articulating an Uncompromising Forgiveness," 537.
51 Smith, "Responsibility for Attitudes," 266.
52 I hold the notions of blame and resentment separate, not because they are completely independent phenomena, resentment can be a viable account of blame, see, e.g. Menges, *Moralische Vorwürfe*; but because *blame* is a more ambiguous term and can be used to refer to different phenomena, such as an attitude, but also the speech-act, see, e.g. Fricker, "What's the Point of Blame?"; or attribution of a causal role, see, e.g. Hart, "Legal Responsibility and Excuses."

5 Accountability for Emotions

In the last chapter, I discussed whether we are in principle ever accountable for our emotions. In this chapter, I focus on the conditions under which it is fair to blame someone for an unfair emotion. While the possibility that we are to blame for an emotion is required for the concept of an unfair emotion, it is not the case that we are always blameworthy for an unfair emotion, nor is it the case that an emotion that we are not to blame for cannot be unfair. The relation is not to be understood as a necessary condition for an individual emotion to be unfair, but as a general condition for emotions being the sort of thing that can be unfair. If emotions were in principle not a thing that we can be blamed for, then calling any emotion unfair would seem somewhat meaningless. It is possible that sometimes an unfair emotion can be blamed on the subject, which is required for emotions to be sometimes unfair.

I have argued that there are ways of holding people accountable that are inappropriate in the case of emotions. We should not sanction or punish people for unfair emotions, and we plausibly don't have any moral duties to feel or not feel certain emotions, since we lack the relevant kind of control over them. But there is a way of holding people accountable that does apply to emotions, namely, the type of relational accountability described in the *reactive attitudes* (RA) literature.[1] According to RA theory, we hold other people accountable for their actions by being disposed to react to their actions with reactive emotions, such as anger, disappointment, hurt, joy, or gratitude. This notion of accountability is relational because it focuses on features of the relation between the blamer and the person being blamed, and not merely on features of the offender.

To establish that emotions can be unfair, I only need to show that we are sometimes negatively accountable for an unfair emotion. However, the upshot of this discussion is more far-reaching. I argue that it can be appropriate to respond to another person's emotions with reactive attitudes. For example, it can be appropriate to resent someone for feeling pitying you or feel guilty for your own envy of your colleague's success. This also includes

DOI: 10.4324/9781003498032-5

cases, for example, of you reacting with gratitude to someone feeling pleasure or enjoyment when listening to a musical performance of yours. This is just as much a case of holding the happy audience member accountable, in a positive way, for their emotions of delight and enjoyment as resenting someone for their unfair fear is a negative way of holding them accountable. All these emotions, of anger, resentment, guilt, but also gratitude count as reactive emotions because they function as a way of holding others accountable for their behaviour or attitudes. Sadly, however, my focus will remain on the negative cases of accountability, and more specifically on resentment, anger, and indignation.

If we apply the RA account of moral accountability to emotions, it follows that to hold someone accountable for an unfair emotion would amount to being disposed to feeling anger, resentment, or indignation in response. For example, imagine you trip and fall, scraping your knee, and your friend simply bursts out laughing. If you held your friend negatively accountable for being amused, you would be disposed to feel angry with your friend for such a blatant display of disregard for your well-being. In line with this basic idea of RA theory, I propose the following account of accountability for emotions:

1. To hold someone morally accountable centrally involves holding them to the expectation that they have an attitude of basic human regard towards others.
2. To hold someone to such an expectation of basic human regard means being disposed to feeling certain reactive emotions, like anger, indignation or resentment, in response to an action or attitude that reflects a defection from an attitude of basic human regard towards you or another person.
3. Since emotions are the kind of mental state that can reflect other attitudes – such as care or regard for others or the lack thereof – they can be subject to this type of holding accountable.

In this chapter, I aim to defend the conclusion that emotions can be the proper target of reactive emotions, and that thereby, we can intelligibly hold people accountable for their emotions. I argue that the RA notion of moral accountability applies to emotions as well as actions.[2]

In Section 5.1, I discuss the relevant notion of holding someone to an expectation mentioned in (1) and (2). This notion serves two roles in RA theory. First, it explicates the idea of holding someone accountable. Second, it provides the conditions under which reactive emotions like anger or resentment count as a case of blaming someone.

In Section 5.2, I discuss the observation that some instances of reactive emotions seem to be appropriate despite not being responses to moral failings. This opens up the possibility of non-moral cases of holding someone accountable according to RA theory. This can pose a problem for my thesis if all cases of appropriate reactive attitudes in response to unfair emotions turn out to be non-moral. I introduce two possible accounts of this distinction, one by Wallace and one by Strawson, neither of which provides a fully satisfying account that applies to unfair emotions. While Wallace's account excludes the possibility of holding people morally accountable for unfair emotions, Strawson's only provides a formal criterion that is not conclusive on the question.

In Section 5.3, I develop my own solution on the basis of Callard's notion of co-valuation. I propose that the relevant cases of holding morally accountable can be captured by the notion of co-valuational expectations that are general enough to apply to all moral agents. In Section 5.3.1, I then go on to outline a rough account of what those expectations involve. I conclude that the expectations that can be defended against any moral agent are expectations of basic human regard. An unfair emotion that reflects a violation of these expectations of basic human regard would count as a proper basis for genuine moral reactive attitudes.

5.1 Holding Others to Expectations

To hold someone to an expectation, in the above sense, means being disposed to feeling certain reactive emotions, like anger, indignation or resentment, in response to actions, attitudes or emotions that reflect a violation of the expectation. There are many possible ways in which someone can be disposed to feel anger or resentment. For example, you could have an anger issue and so be disposed to rage whenever you see the colour red. However, this does not seem like the right kind of disposition to account for the reactive attitudes we feel when we hold others accountable. It is not a disposition to respond to morally objectionable behaviour. Nor is it a response we can intelligibly identify with the notion of blame. What then is this disposition to feel reactive emotions, and how should we understand it?

Wallace formulates the idea that to hold someone accountable is to be disposed to feel, or would find it appropriate to feel, reactive emotions towards them on the basis of your *normative expectations*.[3] For example, I expect you to respect my property, so when you steal my umbrella I am prone to react with anger. However, not all expectations are of this sort. If I know that you get a coffee every day at the same time, I expect that you will get one at that time today. But when you don't, I am not disposed to get angry. The difference between these two kinds of expectation is that the first is a *normative* expectation, while the latter is purely a

predictive expectation. A predictive expectation merely consists in thinking that something will happen, for example, that someone will act in a certain way, mostly due to prior experience or by some other predictive method. Normative expectations, on the other hand, are distinct in that you don't necessarily think that someone will behave in a certain way, but that someone ought to behave in such a way.[4]

Blame is connected to such normative expectations in the following way: A blames B for X if A is either disposed to feeling or judges it to be fitting to feel anger or indignation towards B on the basis that X violates a normative expectation A holds B to. Hence, blame is a special instance of holding someone accountable, namely, under the condition that a relevant expectation has already been violated. Blame therefore is a type of emotional disposition that centrally involves reactive emotions like anger and indignation. I diverge from Wallace's description here and will use the term *resentment* for this emotional disposition, whereas he uses the term to label one of the emotional episodes it disposes us to feel. However, I will continue to refer to both resentment as an emotional disposition and the emotional episodes of anger or indignation as reactive emotions in the following. I further elaborate my distinction between *anger*, *indignation*, and *resentment* in Chapter 6, Section 6.1.

5.1.1 The Reactive Emotions

Which are the reactive emotions that are part of holding someone accountable, like anger, indignation or resentment? And how can we differentiate them from non-reactive emotions, like sadness or contempt, which we might also feel in similar situations? Contempt, for example, also seems to evaluate something about another person as morally objectionable. But this type of evaluation does not respond to a wrongdoing, but to the *bad-being*[5] of another person. Similarly, sadness might be a fitting response to the same act as anger, but it evaluates it as a loss instead of as a wronging. Some losses can be losses of a moral good and therefore by themselves morally bad, and in such a case, sadness might even respond to the same moral fact as anger, but in a different way. But neither of those evaluations are *reactive* in the rather technical sense of the term used in RA theory.

For one, the difference between the reactive and non-reactive emotions is that they target a specific type of moral quality. Contempt may still be an appropriate response to some types of moral failings, such as a vicious character trait, weakness of will, or similar ways of someone being morally bad. Other emotions like sadness might also be about, for example, the loss of a moral good. But neither type of emotion is much an inherently fitting reaction to the moral badness of wrongings as anger, indignation, and resentment are.

The differences between emotions like contempt and the reactive emotions are also apparent in the normative upshot that follows from them. Reactive emotions can be understood as a form of addressing the target and are often accompanied by certain further expectations towards the offender to apologize, feel guilty, or recognize their wronging. In contrast, we would not expect a contemptible person to apologize for being so, or to forgive them in response. Neither does everything that evokes sadness also call for an apology. This suggests that it is simply certain types of emotion, like anger and indignation, that count as reactive versus the non-reactive ones like contempt or sadness.[6]

At first glance, it might seem that unfair emotions would be a more proper target of non-reactive attitudes like contempt. Are emotions not more a matter of how we are than what we do? They appear closer to character traits or affective dispositions than to actions and overt behaviour. As such, emotions would be ideal targets of contempt, and there are many cases in which I agree that contempt for someone is an appropriate response. If someone's emotions are an expression of a moral vice, contempt for that person for those emotions would be perfectly fitting. But just because one moral emotion – like contempt – is sometimes an appropriate response to someone's emotions does not mean that it is the *only* appropriate response. Another moral emotion – like anger – could be an appropriate response as well. Analogously, it can be very much appropriate to respond to an immoral action with contempt for the vicious character trait it reflects, while also responding with anger to the moral offensiveness of the action. Contempt for a bad character trait does not preclude that anger about the moral wrongness is another appropriate response.

5.2 Moral Versus Non-Moral

Many authors writing about reactive attitudes theory distinguish moral from non-moral – or *personal* – reactive attitudes. The distinction reflects that we can experience reactive emotions, like anger, in response to someone failing an expectation that is not distinctively moral. For example, you might get angry at a player for failing your expectation towards them to win the championship final, or the expectation towards your lunch partner to conform with the rules of etiquette. While these reactions seem innocuous, they present a case of reactive emotions that are nonetheless not responding to genuine moral offences. Hence, there seems to be a meaningful distinction between genuinely moral reactive attitudes, which are responses to failing moral expectations, and non-moral reactive attitudes which you feel when someone fails to live up to a personal expectation. In this case, all of these expectations are normative expectations, but they

differ in whether they are *moral* normative expectations or *non-moral* normative expectations.

This distinction is relevant to the question whether we can justifiably hold people morally accountable for their emotions. It might be the case that, while we can hold people to expectations and be disposed to reactive emotions when they fail them, we can only do so in a personal, non-moral sense. This would mean that we cannot morally blame anyone for an emotion and therefore draws into question whether supposedly unfair emotions are actually ever morally blameworthy, or rather merely a personal issue. To find out whether emotions can ever be an appropriate basis for moral blame, we need to get a clearer account of the distinction between moral and non-moral reactive attitudes. I first examine Wallace's account of the moral – non-moral – distinction and show that it runs into trouble when we try to apply it to unfair emotions. I then present an alternative account of the distinction, proposed by Strawson, which gets around the problems of the first one.

5.2.1 Wallace and Moral Obligation

Wallace explicitly identifies the type of expectations relevant for moral reactive attitudes as expecting someone to adhere to moral obligations.[7] Hence, to hold someone morally accountable, according to Wallace, means to hold them to moral obligations. And to hold someone to a moral obligation means to be disposed to feel anger or indignation when this obligation is violated, or judge those emotions appropriately.

As I argued in Section 3.4, it is not very plausible to assume that we have moral obligations or duties to feel or not feel certain emotions. This is because having an obligation, it is commonly assumed, requires that you have a substantial degree of voluntary control over what you are obliged to do. But, as discussed in Section 4.3, we do not have very extensive voluntary control over our emotions. Does this put an end to the idea that we can legitimately be held morally accountable for unfair emotions? I think such a conclusion would be too quick. There are distinctively moral expectations, which might not count as obligations, that still seem to justify moral reactive emotions. If we drop the requirement that an expectation has to be narrowly about a moral obligation to warrant moral reactive attitudes, we can still be morally accountable for unfair emotions.

First, holding people to moral obligations does not explain positive forms of accountability. We don't only blame people for their bad deeds, we also praise people for going above and beyond duty, or admire them for their extraordinary courage or selflessness. An account of the reactive attitudes should be able to account for these positive responses just as well as for negative ones. Positive reactive emotions could be incorporated into an

obligations-based RA account by construing them as reactions to someone acting according to their obligations. However, that seems to overshoot the goal. Someone acting according to their moral duties is often the minimum to be expected. For example, it seems out of place to react with gratitude or admiration to someone passing you by in the street simply because they don't try to mug you. This especially does not account for the positive emotions we might feel towards someone for helping us when they weren't obligated. At a minimum, the account of holding people to moral obligations has to be amended to include reacting to acts that aim at achieving supererogatory moral goods.

The second problem for the obligations-based account is that tying reactive emotions to obligations reintroduces a strong voluntary control condition, as I have argued in Section 3.4. It seems that to be held to obligations requires being able to intentionally follow those obligations. But not all cases where we praise or blame people are cases in which they do things intentionally. For example, we hold people positively accountable for having good ideas. If someone comes up with a spontaneous and genius idea that could save lives, we tend to praise them for it. However, these are just the types of things that we cannot control. In many cases, good ideas are unpredictable and seem to come out of nowhere, even to the person who got the idea. We can try to willingly get a good idea, but it is not typically something we can intend and succeed in doing. We therefore should either stop praising people for their good ideas or acknowledge that holding people to obligations is not the main sense in which we hold people accountable.

5.2.2 Strawson and Quality of Will

Peter F. Strawson, who coined the term *reactive attitudes*, did not formulate an obligation-based requirement for reactive attitudes.[8] Rather, he proposed that what reactive attitudes respond to is a certain quality of will in the other person. For example, if we react with anger to someone lying to us, we react to the perceived malevolence or *ill will* they show us by lying. We don't necessarily respond with resentment to lying as a breach of obligation, but only when we perceive it as expressing an offensive attitude, such as malevolence or disregard for our well-being or autonomy. For example, if you lied to me to distract me from the preparations for a surprise birthday party – assuming you know that I love surprise parties – I would not have any reason to resent you, even if lying were a breach of obligation. This is because you didn't lie out of ill will, but out of friendship or goodwill towards me. However, if you lied to me about the availability of a job opening I was hoping for, just to apply to it yourself, I would have ample reason to resent you. This is because your lie would reflect

an attitude either of disrespect or even of malevolence towards me. Such attitudes of ill will towards someone can be reflected by things other than actions, namely, also by desires, beliefs, and emotions. Instead of expecting compliance with specific obligations, under this alternative account, the relevant moral expectations are directed at certain good or bad attitudes towards other people.

For an emotion to reflect such an attitude of good or bad will does not require voluntary control over the emotion – in clear contrast to compliance with moral obligations. This avoids the problem with Wallace's account, which would require voluntary control and therefore render moral accountability for emotions impossible. What is required is for the attitude to be necessary to understand the emotion – for the emotion to be intelligible. For example, if I am saddened by my local football team winning the national championship, you would probably require additional information to make sense of my emotional reaction. The unhappiness could be due to my attitude of being a fan of the other team. Alternatively, I might have bet a large sum of money against my local team and now lost it all. In the latter case, my unhappiness takes on a different meaning compared to the former case. In either case, for my displeasure to reflect my underlying attitudes, I don't need to have any voluntary control over my emotions, but only the type of rational control presented in Section 4.3.3.

However, on Strawson's account, not all reactive attitudes are moral either. In contrast to Wallace, Strawson distinguishes the moral from the non-moral reactive attitudes by whether they are *generalizable* or not. That is to say, if the emotion is subject-specific, then it is not a moral one, but rather a personal reactive emotion. The kinds of expectations we hold others to, and which give rise to these personal reactive emotions, vary with the type of relationship we have with them. But if we can generalize these expectations to also hold for any other moral agent, even one who doesn't share a specific relationship with us, and feel indignation vicariously – on their behalf – then the expectations involved count as moral.[9] The notions of generalizability and vicariousness are therefore closely related for Strawson. We only feel resentment on our own behalf for personal, subject-specific expectations, for example, when someone maligned your favourite romance novel. It wouldn't make sense to feel the same kind of resentment if they instead bad-mouthed the book of a stranger on the bus. But if the expectation is generalizable, for example when someone verbally attacked a stranger on the bus, then it is perfectly adequate to feel resentment vicariously, on the stranger's behalf.

So far, nothing in Strawson's approach seems to be incompatible with resenting someone for an unfair emotion. Emotions can reflect attitudes of good or ill will, and in cases where this violates a generalizable standard, resentment can constitute a response of moral blame. However, Strawson

only provides this formal criterion that reactive attitudes count as moral when the violated expectations are generalizable. He does not provide any substantive account about what kinds of expectations are generalizable. To fill in this gap and see whether there are generalizable expectations that apply to emotions, I need to develop such a substantive account. In Section 5.3, I begin with Strawson's approach and start with more personal expectations about what emotions other people feel towards us, and then show that we can generalize some of those expectations so that they hold for every human moral agent. I first look at more personal reactive attitudes and what kind of expectations about what others feel we can have in specific relationships, like feeling hurt at a loved one's indifference, or a parent being disappointed at our best try. Then, I broaden the view to the kinds of emotional reactions we can expect any responsible human agent not to feel towards us, like the enjoyment of an innocent's suffering or racist fears.

5.3 Co-Valuational Expectations

If it is not only breaches of obligations that reactive emotions respond to, then which expectations are the ones relevant for resentment and indignation? Strawson's own account doesn't give us much guidance, since it only identifies the relevant moral reactive attitudes as those that we can also feel on behalf of all others. This only provides us with the formal requirement that any moral expectations should be generalizable to hold for all relations between fellow moral beings. But how do we identify this class of expectations – can we formulate a more substantial requirement?

We can distinguish between emotional reactions to a person's actions where we expect them to share our adherence to a norm and reactions where we don't hold such an expectation. For example, you might aesthetically disapprove of someone's taste in food or living room furniture but have no problem tolerating these differences in taste. In contrast, if you disapprove of someone's use of corporal punishment in raising their children, you might not take such a tolerant view, but very much expect them to change their attitudes towards inflicting pain on children. The difference is not necessarily in how you act towards the offender but in the emotional reaction to them. In the former case, you might feel irritated or even a bit disgusted, but you wouldn't resent someone for their culinary idiosyncrasies. In the latter case, however, you might very much feel resentment and indignation.

To get a better understanding of this difference between which norm violations elicit reactive attitudes and which ones we can tolerate, it is helpful to turn to the concept of *co-valuation*, introduced by Callard.[10] Two people co-value something if they not only both value the same thing

independently, but when they both also value the same thing together as a shared project. In such a case, both people care that the other person also values the same thing, and come to expect that the other person continues to uphold their shared value. This co-valuing then becomes part of what constitutes their relationship. Co-valuation further entails that we feel reactive emotions in response to a violation of a co-valuational expectation, since it has become a relevant part of the relation to the offender. In effect, someone's norm violation only evokes your anger when that violation signifies a defection from a co-valuational expectation that you hold them to – as a shared norm. Such a defection means that the other person no longer values something they should value and act or feel out of a disregard for that value – or out of an ill quality of will with respect to that value. The following example illustrates this point:

> Amelia and Ophelia are colleagues who work for the same company. Ophelia has the – what some might call old-fashioned – view that co-workers should cooperate and show solidarity towards each-other. Amelia, on the other hand, has a more competitive, 'dog-eat-dog' view of the business world. When a senior management position opens that Ophelia has been working towards for over ten years, the relatively newer Amelia goes for it, using all the dirty tricks to outmanoeuvre Ophelia and gets the promotion. When Ophelia confronts Amelia, angry with her betrayal as a colleague and co-worker, Amelia responds with incredulity. To her, in business it's everyone for themselves and that they don't owe each-other any loyalty. Were their places flipped, Amelia wouldn't resent nor be angry with Ophelia, either.

The example is meant to illustrate that when someone like Ophelia holds herself and her colleagues to a norm of solidarity, she is liable to react with anger and resentment to a defection by Amelia. Amelia, on the other hand, is not liable to respond with reactive emotions to similar behaviour. The upshot here is that it is expectations of such shared norms that work as a basis for reactive emotions. In the next step, I want to illustrate the plausibility that this also holds for mental states, like emotions, and not just for actions.

You share a love for your local football team with your friends. But you find that over the last year, you have become bored and at times even annoyed with football, and even your team. When you confess your lack of enthusiasm to your friends, their first reactions are feelings of anger, hurt and feeling slightly betrayed. This might last only for a short while and, if you are friends for other reasons as well, they will come to accept this shift in your shared interests. However, in this first instance, your perceived defection from a shared value can evoke responses that are clearly

reactive emotions. Similar responses would not seem unusual in cases of belief or desires, for example, where children confess their lack of belief in god to devout parents or when a partner doesn't share the desire to have children. The same can apply not only to the lack of a certain attitude but also to its presence. For example, when your circle of friends holds a strong working-class cultural identity, but you adore the opera, or when you have a sense of humour that your partner finds annoying or childish.

These cases can all be accounted for by Callard's notion of co-valuation. The defection from a perceived shared project of valuing something together can feel like a betrayal or an offence, and thereby evoke feelings of anger, hurt and resentment. This defection does not need to involve overt action and can be reflected merely by emotions or other mental attitudes. But not all emotions that intelligibly elicit reactive emotions like anger or resentment are unfair. For one, many of these emotions are not directed at the person getting angry. Therefore, even if some of them might be morally objectionable, they would not be unfair. For another, not all emotions that elicit angry responses are morally objectionable. Some, like the enjoyment of the opera or an odd sense of humour, seem perfectly morally innocuous. Emotions of anger as a response to these cases would fall under Strawson's category of *personal* reactive attitudes because they depend on particular personal relationships and particular expectations within these relationships to make sense.

If not all cases of anger are reactions to moral infringements, then are these personal reactive attitudes ever fitting or justified? At first, we might think they are not. Since defection from such particular expectations, like loving the local football team, does not constitute a real offence, anger would not be a fitting response to it. However, the notion of an offence is not necessarily a purely moral one. It can also serve, in just the way described earlier, as a violation of a shared norm, the defection from which is regarded as a personal offence to one's shared valuing. Since these reactive emotions are personal, not everyone can reasonably share them as a reaction to the perceived offences. What is offended, in the case of the football friends, is the sense of shared identity and shared values. An outsider to the group would not have any reason to be angry with you, as the defector, on your friends' behalf.[11] It would not make sense to such an outsider to hold you to the expectation to share in your friends' shared identity.

Contrast this case with the case of the teenager standing at a ticketing machine having trouble selecting the right ticket, and the person in line just behind them getting angry with impatience even though the teenager clearly has a difficult problem that simply takes time to solve. In this case, the shared value in question is that of general decency towards others in public. We don't need to assume a shared special relationship between the teenager and the angry person to hold them to an expectation of mutual

respect – which in this case would constitute the relevant good quality of will. Since this is an expectation we can reasonably hold towards all responsible members of society, we can be justified in feeling indignation or anger towards the impatient queuer on the teenager's behalf.

This difference between expectations that are limited to people inside the personal relationship and those that aren't, I think, is the most plausible interpretation of Strawson's distinction between moral and non-moral reactive attitudes. Since we cannot reasonably expect every member of society to co-value our local football team, we cannot accept this as a moral expectation. But if we can expect everyone to co-value mutual respect or a basic regard for the life or well-being of others, then we can view this expectation as a moral one. Someone acting out of ill will would then mean that they act in violation of such an expectation and without an attitude of respect or regard for others. In the following, I elaborate on what kinds of expectations are general in this relevant sense, and in what relation they stand to the unfairness of emotions.

5.3.1 Basic Human Regard

Are we ever justified in holding other people morally accountable for nothing but their emotions? Whether we are depends, in large parts, on what expectations we can hold them to and whether we can make a good case that other people should share those expectations. Can we generalize the above-discussed relational expectations, which depend strongly on the type of special relationship the parties share and give rise to a relationship-specific form of accountability, to apply to strangers as well? Are there, so to speak, *general* relational expectations?[12] At first glance, it seems trivial that we can. We could simply expect everyone we meet to share the love of the local football team and resent them if they don't. However, while it is possible to hold such general expectations, it does not follow that they are morally justified. It would clearly not be a justified expectation to hold everyone to the expectation to share one's love for football.

It is important to stress that we can of course hold people accountable in a vicious way. That is to say, we can hold people accountable on the basis of unjustifiable relational expectations. For example, if I expect a colleague to show affection for me or give me preferential treatment, and get angry when they do not, I am holding them accountable in an inappropriate and morally indefensible way. Put differently, I hold them to certain morally dubious expectations and am disposed to a range of reactive emotions in response to them falling short of meeting those expectations.

So if we are interested in when it is morally appropriate to hold someone accountable, we need to limit ourselves to the class of morally justified general expectations. Those that concern things which we could expect

every human agent to co-value and value as a basis for their relations to every other human agent. I will call this class of expectations *basic human regard*.[13] What I mean by basic human regard is not necessarily a set of rules or norms, but attitudes we ought to have towards others and through which we relate to them as morally relevant fellow human beings.

What can we expect every human agent to value? In the philosophical literature and broader moral tradition, we find many proposals and examples of evaluative attitudes expected from any moral agent:

- Some religious traditions, like Christianity, teach that we should have unconditional love for others. A plausible interpretation of this type of commandment is that we ought to have a certain type of positive attitude or good will towards our fellow humans. It might even be morally wrong to not have this positive attitude towards others. The Christian Bible (1 John 3:15) even states that hating one's fellow humans is similar, morally speaking, to intending murder.
- Kant's second formulation of the categorical imperative, often called the *humanity formula*, tells us to act in such a way that we never treat the *humanity* in others as a mere means to an end, but always also as an end in itself.[14] *Humanity* in Kant's work means a set of features that mark someone as distinctly human. These centrally involve the capacity for self-guided actions and rational thinking.[15] One interpretation of the humanity formula is that it demands *respect*[16] of other people's humanity, which primarily means respecting their autonomy. It is this attitude of respect towards the humanity of others that is morally demanded of any agent.
- Hume observes that there is a range of natural virtues which we seem to admire, not only in those close to us but in any person – such as *generosity, humanity, compassion, gratitude*, etc.[17] or *benevolence*.[18] In contrast, we seem to equally disapprove of cruelty and inhumanity in anyone.[19] These are all general attitudes towards others that we closely associate with a morally good or bad person. Hence, we come to expect them from others, otherwise we disapprove of them.
- Scanlon argues that morality requires of us that we hold certain attitudes towards others simply by virtue of being 'fellow human beings'. Such attitudes include taking care not to harm others through our actions, helping others when we can easily do so, not lying or misleading, and similar attitudes of general good will and regard. These attitudes define what he calls the *moral relationship* that we should have toward other rational beings.[20]
- Fricker, in her discussion of *epistemic injustice*,[21] stresses the importance of being seen as a credible source of knowledge and an attitude of epistemic trust towards others as being central to questions of justice. She

> counts unearned disbelief, when it comes about through prejudice, as an injustice. This implies an expectation of an attitude of epistemic trust towards others, absent any good reasons against it.

The details of what basic human regard involves are a matter of normative ethics and a discussion far too extensive to include here. What comes out of the examples above is that the kinds of general expectations that can ground moral accountability for emotions could be identified as general attitudes, or sentiments, towards other people. These are likely to include (1) some form of respect for other people's autonomy, (2) a basic regard for their well-being, and (3) some form of basic trust or initial assumption of other people's reliability and good will. These are all attitudes that we should expect people to have, rather than duties we would expect them to comply with as per Wallace. Hence, the expectation of basic human regard is an expectation towards the quality of someone's underlying attitudes, the quality of their will or the sentiments they hold towards others – and not towards the quality of their overt actions. Therefore, while actions can reflect these underlying attitudes, so can emotions – and possibly other mental states like desires or even beliefs.

We expect people to hold these attitudes of basic human regard not only towards ourselves but also towards third parties. In this respect, these expectations differ from personal expectations, such as those based on shared interests or personal relationships. As a result, we can also be disposed to feel reactive emotions when people defect from basic human regard towards others, not only ourselves. To use Strawson's term, we can feel indignation vicariously, meaning on behalf of others, when someone shows them disrespect, disregard for basic well-being or unjustified mistrust. It is in this sense, that we hold people morally accountable to basic moral norms, and not merely to expectations of personal relationships.

5.3.2 Fitting Reactive Emotions

The question that still remains is, whether reactive emotions like anger, resentment or indignation are appropriate responses when an attitude of disregard is expressed merely in the form of an emotion – and not through overt actions or behaviour. I have claimed in Section 4.4.4 that voluntary control is not a requirement for a reactive attitudes theory of blame and that therefore, unfair emotions are just as much a viable basis for blame as actions. However, I still need to show that this is the case and that an unfair emotion can be the basis for appropriate anger or resentment.

I have argued that there are certain attitudes of basic human regard which we can morally expect from others. These are values we expect others to hold and share as the foundation of the basic relations we all have

with our fellow human beings. In this section, I focus on the connection between these expectations of basic human regard and the reactive emotions primarily relevant to blame. The basic idea is that moral anger and indignation are justified responses when we experience others defecting from these shared values and when their unfair emotions reflect such a lack of basic human regard – an ill quality of will – towards us or others. To get a clearer picture of whether we can justify these reactive emotions in response to a defection from basic human regard, we need to look at these individual reactive emotions. Is anger, for example, a justifiable reaction to some stranger showing disregard for your well-being by laughing at you trip and fall? And would indignation be a similarly appropriate response if the same happened to someone else?

I argued in Chapter 4 that anger and resentment are not primarily forms of sanction or punishment. While they do shift the subject's motivational stance towards the target to a more aggressive or confrontational direction, they are not simply triggers for violent behaviour. Anger is an emotion that changes how you think about someone and your desires and wishes for that person. If it is *pro tanto* morally justified to react with angry behaviour to someone showing you disregard, even if the person is a stranger, then it would also be intrinsically morally justified – and thereby *fair* – to react with anger. And if it is fair to react with anger, then it is also fair to have the disposition to feel anger. As discussed earlier, having such a disposition just means to blame someone, namely, to be disposed to react with a given set of reactive emotions – anger and indignation – to the breach of an expectation. So, if anger is an appropriate response to, for example, amusement that reflects an objectionable disregard towards you, then being disposed to feel anger in response, and thereby blaming someone for being amused in such a way would be fair. Put more formally:

1. To blame A for X just means being disposed to react with anger and indignation to X on the basis that X reflects a violation of a moral expectation.
2. Such an emotional disposition is fair if, and only if, the emotions it disposes to are also fair.
3. Therefore, blaming A for X is fair if, and only if, the anger or indignation it disposes to in response to X – on the basis that X reflects a violation of a moral expectation – are fair.

I have argued for (1) throughout this chapter. And I take (2) to be a plausible extension of the concept of intrinsic moral justification to emotional dispositions. Therefore, to show that the emotional disposition to feel reactive emotions is fair requires showing that the episodes of reactive emotions it disposes to are fair. Furthermore, this type of emotional

disposition counts as a type of justified moral blame, if the episodes of anger or indignation it disposes to are intrinsically morally justified because the unfair emotion reflects a violation of the moral expectation of basic human regard.

What remains to be shown is when anger and indignation are fair, and whether they are fair in cases where a lack of basic human regard is reflected merely by another person's emotions. This might look like a piecemeal approach to establishing that we can be justified in holding others morally accountable for their affective attitudes towards us. But if the same range of reactive attitudes are appropriate in response to unfair emotions, as are appropriate in response to breaches of duty, then we have a solid argument that we are sometimes to blame for our emotions. In the next chapter, I examine anger, indignation and resentment, which are usually regarded as the primary reactive attitudes relevant to blame, and whether they are appropriate responses to unfair emotions. I will address some challenges to my account, namely, (1) whether resentment is not itself a type of ill will and therefore in danger of generating an infinite regress. And (2) whether resentment and the anger episode it disposes to are ever-fitting reactions to mere emotions.

Notes

1 See, e.g. Strawson, "Freedom and Resentment," 1962; Wallace, *Responsibility and the Moral Sentiments*; Macnamara, "Holding Others Responsible"; McKenna, *Conversation and Responsibility*; McGeer, "Scaffolding Agency."
2 This type of accountability might also apply to other mental states, such as beliefs or desires, but I don't go into those questions here.
3 Wallace, *Responsibility and the Moral Sentiments*.
4 Wallace introduces this distinction between type of expectation in Wallace, *Responsibility and the Moral Sentiments*, 20–25, however, he does not use the terminology of *normative* versus *predictive* expectation. I introduce these terms here as a useful shorthand.
5 See Bell, *Hard Feelings*, 16/39.
6 Some authors disagree with this distinction. Bell, *Hard Feelings*, for example, argues to the contrary that contempt can be seen as a reactive attitude as well and shares many of the qualities of anger or resentment. However, contempt is still not about wronging but bad-being and I therefore maintain my distinction.
7 Wallace, *Responsibility and the Moral Sentiments*, 33–40.
8 Strawson, "Freedom and Resentment," 1962.
9 Strawson, "Freedom and Resentment," section 5.
10 Callard, "The Reason to Be Angry Forever," 130.
11 With the exception maybe of fans of others teams that nonetheless share a similar sense of loyalty, even though they don't support the same team. In such a case, we might be moving away from mere personal expectation to a more moralized expectation that if you are a fan, you should stay loyal.
12 This is a general problem for relational accounts of morality. See, e.g. Parfit, "Reasons and Persons"; Singer, "Famine, Affluence, and Morality."

13 There might be good arguments that we should include non-human animals or any sentient being in this set of expectations. However, the affective responses of non-humans might look very different, or they might not even have emotions in any recognizable form. Therefore, the specifics of what we hold non-humans accountable for could vary drastically.
14 Kant, "Grundlegung Zur Metaphysik Der Sitten," vol. 4, 429.
15 See Korsgaard, "Kant's Formula of Humanity."
16 See, e.g. Darwall, "Two Kinds of Respect," for a discussion on the type of respect that best capture Kant's *respect for humanity*.
17 See Hume, *A Treatise of Human Nature*, T 3.3.3.3.
18 Hume, *A Treatise of Human Nature*, T 3.3.3.6.
19 Hume, *A Treatise of Human Nature*, T 3.3.3.8–9.
20 Scanlon, "Moral Dimensions," 140.
21 Fricker, *Epistemic Injustice*.

6 Fair Resentment for Unfair Emotions

So far, I have been arguing under the assumption that people frequently blame each other for unfair emotions, and that this practice can be philosophically vindicated by applying relatively well-substantiated moral theories. However, if the arguments that support this everyday practice had absurd or highly implausible implications, the line of argument presented here could also be used as an informal *reductio*, to show that this everyday practice should be abandoned. Put more bluntly, if it turns out that to justify blaming others for their unfair emotions means that we have to accept absurd theoretical conclusions, it would be better to accept that we should not blame others for their emotions. This might not be an internal inconsistency in my account, but the upshot that my conclusions are so implausible that it is more likely that something about my assumptions must be wrong. In this chapter, I discuss the most troubling conclusion that might lead us to reject the whole account, and show how it can be resolved.

In Chapter 3, I have presented my account of when an emotion is unfair to its target. An emotion is unfair when the emotion's action tendencies pose a moral hazard to the target – a hazard that is not *pro tanto* justified by its fit-making conditions because those conditions are not met. And in Chapter 5, I have shown how the reactive attitudes theory of moral accountability can be applied to emotions. We can hold people accountable for their unfair emotions by being disposed to feel certain reactive emotions, like anger, indignation, and resentment, towards them when the unfair emotion they feel reflects a sentiment of ill will towards us that violates the requirement of basic human regard.

Combining these two accounts raises a difficult challenge. If you resent someone, does that not also constitute a type of ill will you hold towards them? And does not anger itself pose a moral hazard to its target? Anger is traditionally portrayed as an emotion that aims at payback and revenge. But to seek revenge for nothing more than someone feeling in a way you find unfair seems an overreaction at best, and an absurd escalation of aggressiveness or even violence at worst. If this is the case, whenever we

DOI: 10.4324/9781003498032-6

hold someone accountable for an unfair emotion by way of resentment or getting angry with them, we would ourselves become blameworthy for feeling an unfair emotion. The assumption from Chapter 3, that emotions can be unfair, would make it impermissible to hold people accountable for their emotions under the reactive attitudes theory of accountability presented in Chapter 5. If my account were to have such an implication, would that not count against the account, and the practice it seeks to support?

This defeating conclusion can only be avoided if it can be shown that anger and resentment are sometimes fair responses to unfair emotions. The aim of this chapter is to show that this is the case. I argue that if anger and resentment are fair responses to moral offences or ill will, then they are also fair responses to unfair emotions. To do so, I first make some distinctions between how I understand the notions of resentment, anger, and indignation in Section 6.1. I also show how these affective states hang together, that resentment disposes to repeated anger episodes, and how therefore the appropriateness of resentment depends, in part, on the appropriateness of those episodes of anger.

Next, in Section 6.2, I address the challenge that resentment itself constitutes an attitude of ill will, and therefore any emotion that reflects such resentment would count as resentable itself, and so forth, *ad infinitum*. To diffuse this worry, I argue that resentment is, despite initial appearances, compatible with basic human regard, and does not necessarily constitute an attitude of ill will.

In Section 6.3, I address the worry that anger is necessarily a violent emotion that aims at revenge or retribution. If this is the case, there are good reasons to reject that anger is ever a fair response to someone merely feeling an emotion. To reject this conclusion, I propose two viable strategies. Either, in Section 6.3.1, that the desire for retribution or revenge inherent to anger is less objectionable than it might first appear; or, in Section 6.3.2, that anger does not in fact aim at revenge, but rather is a way of protesting a moral wrong and aims at recognition for the imposed wrong.

6.1 Distinctions

Before delving into the specific problems resentment, anger, and indignation may pose for my account, I have to clarify how I understand and differentiate these attitudes. I will broadly follow Deonna and Teroni's[1] distinction between *emotional episodes* and *emotional dispositions* to distinguish anger and indignation as emotions on the one hand, and resentment as a disposition on the other hand. But I remain agnostic about which phenomenon best captures the notion of emotion itself – whether by 'an emotion' we commonly mean emotional episodes or emotional dispositions or both.

I take anger and indignation to be short-term, felt *emotional episodes* directed at a specific target, like a person, a group, an object, or even a state of affairs. They are typically felt, involving bodily sensations, and have a characteristic motivational profile. When angry, we relate to the target as a frustration to our goals and are motivated to more forcefully or aggressively confront it and get rid of the problem.

I understand resentment as a second-personal *sentiment*, an *emotional disposition*. Resentment can be much more long term than individual episodes of anger and, as an emotional disposition, disposes the subject to such episodes. In contrast to episodes of anger, resentment is only felt when it manifests in an emotional episode – such as, typically, anger or indignation. But just like anger, resentment is typically directed at a person or group of people and involves a specific wrong done to you or other people on whose behalf you resent. Resentment is therefore different from anger, which is a more basic emotional episode that can be appropriately directed at anything that seems offensive, even without knowing its author. But it is also second-personal, in contrast to other moral emotions such as contempt, for example, which is a third-personal emotion.

These affective attitudes also have different appropriateness conditions. To illustrate these distinctions, consider the following scenarios:

> At a party, a fellow party-goer has put out their cigarette in your cup of beer standing on the small balcony table next to you.

In this first scenario, it seems fitting to get angry at the party-goer and confront them about what they did. It also will be fitting to stop being angry just as quickly if it turns out to have been an accident, and they meant to put the cigarette in their own cup just next to yours. It would have been just as fitting to get angry about someone putting out their cigarette in your beer if you did not know who it was. The situation itself is worthy of a slightly angry response. In either case, anger shows itself to be a second-personal emotional response since it responds to someone acting badly towards *you*, by putting out their cigarette in your drink. Even if it was your friend's cup, you would still get angry vicariously, on the behalf of your friend. In either case, the wrongful behaviour and the response have clear agents and victims or targets.

To contrast this second-personal moral response, of reacting to an actual identifiable offence, with one that only involves a judgement about someone's character, consider the following variation:

> At a party, a party guest puts out their cigarettes carelessly in the next best cup wherever they happen to finish one.

In this second case, it seems appropriate to judge the guest as careless and maybe feel a little bit of contempt for them. If the party guest hasn't so far damaged anything or ruined anyone's drink, you have no reason to get angry at them. There is, so-far, no identifiable victim of their behaviour. However, their general lack of concern for other party-goers reflects a generally bad attitude or lack of moral character, and to feel contempt on that basis seems fitting. This type of third-personal response is more about evaluating the quality of the party guest's character rather than a type of blame for any specific action.

Finally, to contrast responding with anger to responding with full resentment, consider the third version:

> At a party, Dave, who has been out to get you all evening, puts out his cigarette in your beer, all the while looking at you with an expression of contempt.

In this case, Dave's intentions and motives are rather clear – and they are personal. Not only is it appropriate to be angry with Dave, his action of ruining your beer is a clear expression of his hate for you, and resenting him for it would be appropriate. For one, this case of resentment is also a second-personal response since there is a clear agent, Dave, a wrongful behaviour, and person who is wronged by it, you. Your resentment in response mirrors this by resenting Dave for the overt wrong done to you. In contrast to anger, the resentment in this case seems appropriate due to the high degree of certainty about the objectionable motives of the offender. That is not to say that anger requires being unsure about the target's motives, merely that it does not require it so much. In effect, justified resentment requires a high degree of certainty, meaning a defensible case that the offender's conduct reflects a lack of basic regard for you.

Under these conditions, it becomes clear why it is justified to be disposed to feeling anger again when confronted with the offence of the offender. There is a clear difference to the first situation, where your anger can easily be resolved by the revelation that it was an accident and that the perpetrator did not mean to offend you. In this scenario, Dave's disregard for you is only expressed by his actions, but not identical to the actions. Hence, merely resolving your anger about the ruined drink does not get rid of the real issue in the room – namely – that Dave has no regard for you. The issue therefore remains open and unresolved, and the offender maintains an attitude towards the subject that they find morally intolerable. Under such circumstances, as long as Dave does not have a change of heart, and stands behind the expressive meaning of his actions, it remains fitting to be disposed to feel anger over it, since it remains an offence to you.

Resentment is just such an emotional disposition that typically disposes to episodes of anger or indignation.

The reason why I focus on relatively trivial wrongs here is to already highlight the importance of motive for the appropriateness of the responses. I argue that even when we leave out any material consequence, like a ruined drink, and replace it only with emotions expressive of those motives, the appropriateness conditions of reactive attitudes remain equally satisfied. For example, if we replace cigarettes in drinks with rolling their eyes, a typical expression of annoyance, we get similar results.

On the one hand, if you ask your flatmate to bring out the rubbish after the party, and they roll their eyes, that can be slightly angering, but no reason to resent them, unless you know that it is an expression of some larger unresolved issue they have with you. On the other hand, if you know your flatmate regularly gets annoyed when people ask them to do anything, that is a reason to judge them on their character, but not to resent them.

6.1.1 Resentment

Resentment, in contrast to anger, is a more stable and often more long-term affective state towards a person or group of people. I take resentment to be a long-term emotional disposition towards a person. Emotional dispositions of this kind can further be divided into single-track dispositions – which only dispose the subject to a single type of emotion – and multi-track dispositions – which dispose the subject to several different types of emotional episodes, depending on the situation.[2] Resentment certainly seems to involve the disposition to repeatedly experience certain emotions, like anger, when one is again confronted with the offence. Under this aspect, resentment could simply be a disposition to feel angry under the specific conditions that one views the target as a moral offender.

We could understand resentment as a single-track emotional disposition if we take anger to be the only emotion it disposes one to. This would be analogous to fearing the neighbour's dog. There are two senses in which you can fear the neighbour's dog, either by feeling an episode of fear when confronted with the dog or having a general disposition to feel fear when confronted with it or even at the mere thought of the dog. In the first sense, you experience an emotional episode, in the second sense, you have a single-track emotional disposition towards the dog. In this sense, to resent someone would simply be the disposition to feel anger towards that person.

Alternatively, we could understand resentment as a multi-track disposition that disposes a subject to more than just one type of emotion. For example, love for someone disposes a subject to feel many different emotions, like affection in their presence, joy at their success, or fear for their safety. Similarly, resentment might dispose a subject not only to feel anger

with the target, but also, for example, satisfaction at seeing the target face justice. It is hard to tell whether resentment disposes one to feeling this sense of satisfaction in the same way as with anger, or whether it should be considered more as an incidental side effect. Whether or not resentment disposes a subject to feel other emotions, the central claim here is that it is an emotional disposition that centrally disposes a subject to feel anger at a target for some offence or injustice.

As an emotional disposition, resentment is not directly felt, but only via the anger and maybe other emotional episodes it disposes one to feel. It also inherits much of its motivational relevance from anger, since it influences the subject's behaviour through the emotions it disposes them to feel. I therefore return to the discussion of the relation between the appropriateness of anger and the appropriateness of resentment in Section 6.2.

6.1.2 Anger and Indignation

It seems that anger and indignation are the same in many respects, for example, they share many of their felt qualities and action tendencies, which suggest that they are likely the same type of emotion. However, indignation can be distinguished from other forms of anger by its moral character and potential to be felt vicariously. For example, we might react with indignation to a politician taking bribes but would simply call it 'anger' when we react with a similar emotion when the person in a queue in front of us takes an unreasonably long time. Hence, it is likely a type of anger that we associate with responding to violations of moral norms or duties. Indignation then has a distinct meaning not because it is a distinct type of emotion, but because of its elicitation conditions.

While anger and indignation can be distinguished by their second- and third-personal perspectives,[3] I will treat them broadly as the same type of emotion.[4] This seems plausible when we consider that we also often call an emotion 'anger' in both moral and non-moral circumstances. For example, you can get angry with your computer or your car for not working as desired, which are both instances where your anger does not imply perceiving the machine's failure as a moral wrong.[5] You can also get angry with your friend for lying to you, in which case you very much view it as a moral infraction. In contrast, we don't seem to call the angry emotions you feel about your computer indignation. But it does not seem unusual to call your anger with your friend for lying 'indignation' as well as 'anger'. In this sense, indignation seems a narrower type of moral anger.

We also tend to call anger which we feel on other people's behalf 'indignation'. This makes sense if we accept Strawson's theoretical distinction between moral and non-moral reactive emotions,[6] which I have presented

in Chapter 5. According to this line of argument, what makes a reactive emotion moral just is that we can justifiably feel it vicariously, meaning on other people's behalf. This suggests two possible theoretical demarcations between anger and indignation. We could either call any case of third-personal moral anger 'indignation', which we feel vicariously, or we could simply call any case of moral anger 'indignation', which we can potentially feel vicariously but also on our own behalf.

Following Strawson's usage, I will be content with calling all moral anger indignation but acknowledge that the paradigmatic use of 'indignation' is the feeling of anger at a moral wrong done to others. Since the third-personal case also includes the restriction that legitimate anger on someone else's behalf is limited to genuinely moral anger, I will focus on such cases in the following discussion.

6.2 Resentment as Ill Will

Resentment, understood as a dispositional state, has the potential to affects its target it some clearly negative ways. As described earlier, it disposes the resenter to anger towards the target, but also likely other attitudes that seem to express bad will or ill intent towards the target of the resentment. As such, resentment for unfair emotions faces to urgent problems. First, it seems that, as itself a form of ill will-based emotional state, any resentment would again warrant the same type of negative response, escalating a circle of negative reactive attitudes.

Second, as a dispositional state that has potentially severe consequences for other people, to justify resentment, we depend on an assessment of the target's own quality of will. But knowing the state of mind of any other person is a notoriously hard thing to do, and in cases of a potentially severe response, we require an equally high degree of certainty to justify it. This leads us to the uncertainty problem: that mental states, like the quality of will of others, are notoriously hard to be certain of.

If resentment is a negative attitude towards another person and disposes the subject to feel anger towards the target of resentment, as I have suggested in Section 6.1.1, doesn't that anger then reflect an attitude of ill will or disregard? If so, an instance of holding someone accountable by resenting and feeling anger towards them could itself be worthy of blame and resentment, and so on, *ad infinitum*. While this would not amount to a strict logical inconsistency in my account, it would be a rather undesirable consequence. The upshot would be that we can blame people for their unfair emotions by resenting them, but not without making ourselves into a legitimate target of blame, and so on. To avoid this consequence, we need to establish that resentment is at least not always a type of ill will that itself is a valid target for reactive attitudes.

First and foremost, it would be implausible to claim that resentment is not a negative attitude towards a person. This admission already makes resentment look like a type of ill will or disregard. Furthermore, if I yelled at you out of resentment for you, it seems plausible that this would be sufficient grounds for you responding with reactive attitudes, such as anger or more resentment. According to the reactive attitude theory,[7] you would then seem to be justified to react in such a way, since my yelling at you would reflect an attitude of disregard towards you. Resentment would then seem to be a suitable ground for responding with reactive attitudes in return. And if resentment were sufficient grounds for reactive attitudes, it would likely be so because it is an attitude of disregard itself.

However, while examples like this make it seem like that is the case, we have to reconsider this first appearance. Once we examine the case in more detail, it will become apparent that it is not the resentment itself that amounts to ill will, but rather the context that determines whether resentment reflects a lack of basic regard or not. In the following, I argue that resentment is not incompatible with basic human regard, and therefore does not automatically render the resenter worthy of resentment themselves.

Whether resentment would provide the basis for someone else's resentment in return depends, as I have argued for other affective attitudes, on how it is motivated – on whether it is a misguided sentiment or simply misplaced. For example, I could resent you because I mistakenly believe that you stole my umbrella and are unapologetic about it, because you never apologized and greeted me in the hallway as if you did nothing wrong. This would be a clear case of mistaken resentment, since the initial reason for my resentment would be based on a false belief, because you didn't steal my umbrella. But the certainty that you stole my umbrella intentionally and unapologetically would be reinforced by your nonchalant behaviour afterwards. Hence, my resentment would be warranted from my perspective, but nonetheless unfitting because it is misplaced.

This type of mistake does not qualify as a lack of basic human regard, and therefore does not justify you resenting me in return. The notion of basic human regard I have discussed in Section 5.3.1 does not forbid any and all negative attitudes towards others but only demands maintaining a minimum of respect, regard for their well-being, and unprejudiced initial basic trust. For resentment to be incompatible with these demands, it would either need to involve a failure to respect someone's basic autonomy, a desire to harm them, or be inherently prejudiced. None of those seem plausible.

On the first two points, claiming that resentment necessarily involves disrespecting the target's autonomy and disregard for their well-being, would be a misrepresentation. This should become clear when we consider that you can resent a friend or loved one without desiring them to

needlessly suffer or disrespect their autonomy. To the contrary, resentment often even involves the desire that the target come to acknowledge their wrongdoing and regret it, which is only genuine if it happens out of their own autonomous reasoning and reflection. I nevertheless maintain that resentment disposes a subject to feel emotions of anger and indignation towards the target – and anger is often claimed to involve desires to harm or degrade others. Against this assumption, I argue in Section 6.3 that anger does not bluntly aim at inflicting harm or suffering on its target. If I am correct, then even if resentment disposed the subject to feel anger, that would not be incompatible with basic regard for the target's well-being.

On the third point, while resentment can be based on other prejudiced attitudes, it is not itself inherently prejudiced. This is because it is already based on conditions. Resentment is based on the perception that a prior incident was an offence that reflected the offender's disregard towards you. Hence, even if resentment involves mistrust of the target, that mistrust is not necessarily unconditional or prejudiced. Basic human regard is only incompatible with mistrusting others from the start and without cause. This point should become clear when we compare the above example with one where resentment is in fact based on a prejudicial attitude.

A more troubling possibility would be that I resent you out of a vicious or malicious attitude towards you. For example, I could consider myself your social superior and hold you to a norm of deference towards me, that you always bow to me and get out of my way whenever I walk by, due to my morally indefensible world view about inherited social class. In such a case, I would resent you for not showing the due submissive and deferential behaviour of making way and bowing in my presence. I would not be mistaken about any descriptive fact, you actually did not adhere to my expectations and did behave as you would towards any other social equal. This type of resentment would again be inappropriate in some way, but other than the misplaced resentment above, in this case resentment would be misguided and reflect an indefensible attitude towards you.

This second case of resentment is very much worthy of resentment in return. An attitude as described earlier can conflict with basic human regard in several ways. First, deference could mean that one person is expected to forgo some degree of well-being in favour of the other person's well-being. Second, deference seems incompatible with accepting a person's full autonomy and accepting them as an epistemic equal. It conflicts with a person's autonomy because it presumes that in some matters, the deferential person has to submit to the socially superior's will without good reason. Third, it also conflicts with epistemic equality if deference is not only expected in matters of autonomy and well-being but also in factual disagreements or differences in understanding. And all three expectations are held for no reason but relative social role.

While the specifics of a case can become complex very quickly, where the moral quality of resentment depends on further attitudes that might in return depend on other attitudes or beliefs, the point here is merely that while resentment *can* be morally objectionable it is not *necessarily* or inherently an attitude of ill will or disregard. In the best case, resentment is conditional on the target's prior wrongdoing and does not involve desires to hurt or make the target suffer. In the worst case, it might be based on horrible misanthropic attitudes and reflects complete disregard for the target. In neither case, however, does resentment itself constitute an attitude of ill will or disregard. Even in the latter case, where there is ill will involved, it is not due to resentment itself, but due to the context and further underlying attitudes.

6.3 The Aim of Anger

I have suggested that the most viable way of morally blaming someone for an unfair emotion is to resent them for it. For this account to work, resentment has to be morally justifiable as a response to unfair emotions. Both resentment and anger are tied, directly or indirectly, to the motivational profile of anger. Hence, if we are ever to blame someone for an unfair emotion, we need to be able to justify the motivational profile of anger in response to a mere emotion. But if anger aims at nothing but inflicting suffering on others, then the chances of justifying it as such a response seem slim.

So, when is anger a fair emotion towards others? Anger is typically said to be a fitting response to an offence, meaning an unjustified harm, endangerment, or interference with someone's autonomy. However, *offence* can be interpreted as a much broader category than merely including moral offences, and therefore anger could also be fitting as a response to things like inanimate objects or circumstances that interfere with or frustrate the subject's goals. I am not opposed to such a broader interpretation but focus here on the moral cases, where the source of frustration is other people and the interference is constituted by the violation of a moral norm.

In the last chapter, I argued that unfair emotions can constitute such a violation of a moral norm if they reflect underlying attitudes that are incompatible with basic human regard. As a result, these cases of malicious unfair emotions should count as a fitting object for anger and indignation. The following example is meant to illustrate a situation where anger is an appropriate response to someone feeling an emotion that reflects such a vicious underlying attitude:

Try to imagine that you deeply hate your next door neighbour. You hate him so much that, when one day he falls down the stairs and doesn't get up any more, you feel a spiteful joy. To exclude any violation of duty on your part, imagine further that you nevertheless run to aid him, doing your duty

to help someone in a potentially life-threatening state. When your flatmate, who knows of your deep-seated antipathy but does not share it, rushes over to help you in calling an ambulance and stabilizing your neighbour, she angrily whispers to you: 'Wipe that smirk off your face!'

Why would she be angry? You came to your neighbour's aid and did your moral duty. However, her anger is less about what you do than about your attitude towards the situation. Your help is a reluctant one, and you partly enjoy your neighbour's misfortune. That means that there is a part of you that would like nothing more than to sit back and let things play out. The motivational profile of enjoyment is more plausibly about keeping things as they are because they seem just right, and less about changing them for the better.

Let's grant that your enjoyment is morally objectionable. It poses a clear moral hazard to your neighbour in that, if you followed its motivational direction, you would let him lie there and potentially bleed to death. You also have no excuse for it, since you are not misinterpreting the situation as something like a prank or skit. Your enjoyment is therefore not misplaced, but misguided by your underlying hatred.

Is your flatmate's anger an appropriate response here? How does an aggressive or forceful confrontation fit as a response to you violating her expectation of basic human regard? If we follow a traditional view of anger as an emotion aimed at revenge, as promoted by authors like Nussbaum,[8] it seems dubious that anger would be appropriate in this situation. On her account, anger inherently involves a desire for revenge. But this creates a two-pronged problem for the moral appropriateness of anger.

On the one hand, anger could genuinely be about a moral wrong and a response to it. But in this case, anger might be morally laudable for addressing the right problem, but it is vicious because it poses the wrong solution. Since anger aims at payback for the wrong, it merely motivates inflicting some form of harm or ill on the perpetrator. But inflicting more harm or some other form of ill, as a form of payback, does not resolve the original wrong. Therefore, Nussbaum charges that anger involves a certain type of magical thinking. It is a mistake to think that more harm will somehow remedy already inflicted harm.[9]

On the other hand, if anger is not actually about the moral wrong but something else, it might be the right means, but for a morally insidious goal. Nussbaum suggests that your anger might mainly be a reaction to insult and the diminishment of your relative social status that accompanies a moral offence against you. If the aim of anger then is the downgrading of the offender as payback, this might work as a remedy, but not a morally defensible one. Since, in this case, anger is no longer concerned with making the world right again and remedying wrongs, but mere obsession with relative social status.[10]

There are at least two strategies to salvage the claim that anger is an appropriate response to your unfair joy. First, we could accept, against its initial implausibility, that a desire for revenge is an appropriate response for your flatmate to feel towards you. At first glance, this seems the least promising option. However, if we allow for a broad enough conception of revenge, there is an argument that it can nonetheless be fitting to desire a payback in kind in response to an unfair emotion, as I discuss in Section 6.3.1. The second option is to question the claim that anger in inherently tied to a desire for revenge, as traditional accounts assume. Several authors have recently raised doubts about the supposed intimate link between anger and revenge.[11] This second option, while it relies on less developed accounts of anger, nonetheless is in a better position to give a plausible account of instances such as the flatmate's anger over your unfair joy. I discuss such alternative accounts of anger and how they can account for fair anger in response to an unfair emotion in Section 6.3.2.

6.3.1 Revenge and Retribution

If we maintain the view that anger inherently aims at revenge or retribution, the question becomes, under what conditions would it be appropriate to respond with such a desire? First and foremost, we have to assume that a desire for revenge is appropriate in response to anything – not just whether it is an appropriate response to another person's emotions. Nussbaum denies the appropriateness of anger even in the case of wrongful actions.[12] Others are more willing to grant that getting angry at injustice is a justified response. But even if we can defend that a retributive desire is an appropriate response to wrongdoing, it seems excessive as a response to an unfair emotion.

In the case of wrongdoing, for example, if you physically assaulted your neighbour, payback of a similar kind, say, via the imposition of some form of punishment by the state, could maybe still be defended by some instinctive sense of fair payback. But if the only offence you have committed is to feel a certain way, then actual physical violence in retribution seems excessive and not even *pro tanto* justified. In effect, if anger involves such a desire for violent payback, it would constitute a moral hazard and therefore be an unfair emotion according to my own account.

If anger is never justified as a response to mere emotions, even if they are unfair, then we should not hold people accountable for their emotions as described in the last chapter. This is because, as I have argued, to hold someone accountable means being disposed to feeling anger, among other reactive emotions, in response to the other person's violation of basic human regard. But if it is never appropriate to feel anger in this way, it is also not appropriate to be so disposed. If holding someone accountable for

their emotions is in principle never justified, then this raises the question whether someone feeling an unfair emotion is even morally criticizable in the way I have suggested.

Can this collapse of my account be avoided without giving up the notion that anger is inherently tied to revenge? One way of avoiding this consequence is to question the initial appearance, that desiring payback for a mere emotion is excessive. In the case of a wrongful action, like physical assault, desiring payback might mean wanting to inflict similar pain or suffering on the assailant. However, in the case of an unfair emotion, such a response would not be *proportional*. Rather, it would be an excessive escalation of force.

But what would a proportional payback look like in the case of an unfair emotion. If we take the example of schadenfreude at your neighbour's accident, what would count as revenge might not be for your neighbour or your flatmate to cause you harm, but merely to adopt a similar attitude of disregard towards your suffering. This may seem callous, but it would still be in line with a revenge view of anger and avoid the threat of escalating the desire for violence. It could even be argued that proportional revenge can lead to the offender experiencing the kind of wrong they have caused for themselves, and thereby learn what it is like to be on the receiving end. Even if this is not the motivating aim of a desire for revenge, this type of epistemic benefit could offer some external or instrumental justification for the practice as a whole.

Alternatively, the kind of revenge anger motivates could simply be for the offender to feel guilty about their wronging. Bennett argues that the common assumption that retribution[13] is merely the desire to inflict pain is a common misrepresentation of retributivism and that a more nuanced account can help highlight the virtues of a retribution account of blame.

> We withdraw from wrongdoers, cutting them off, no longer treating them as people with whom we share a community and to whom we have special ties and duties. In my view, the behaviour characteristic of moral disapproval expresses the alienation of the offender from our moral community.[14]

Bennett focuses on this notion that the offender has alienated themselves from the rules and values of the society, and the behavioural tendencies of those who blame the offender both reflect this alienation back onto the offender and put pressure on them to re-commit to the rules and values they have violated. However, simply blaming someone neither necessarily involves cutting all special ties to them, nor does it necessarily involve forsaking all special duties towards them. Blame and resentment are much more focused than that. You might not talk to a friend because they

offended you, but still cherish your friendship and desire to maintain it, or help them out when they need a place to crash.

There is a difference between aiming at another person's suffering or aiming at something else that involves another person's suffering.[15] What resentment and anger aim for is an admittance of wrongdoing and a renewed commitment to the broken norms. This might necessarily involve feelings of guilt and remorse, which in turn necessarily feel bad and involve a kind of suffering. But the suffering inherent to guilt is not just any type of suffering. It is a bad feeling due to the recognition of one's own wrongdoings. Therefore, it is unavoidably tied to a genuine acknowledgement of one's own wrongdoing. That anger and resentment aim at something that necessarily involves this specific kind of suffering does not imply that they aim at suffering simply for the sake of suffering. Suffering can be inflicted in many ways, but not all of these satisfy feelings of anger or resentment. If you think you are angry but simply want to see someone suffer, you might be mistaking hate for anger. I think it is not an uncommon phenomenon that people mix up hate with anger or blur the distinctions between them. I come back to this distinction in Section 6.3.2.

Nonetheless, if it is true that retribution aims at a specific type of suffering that is an inherent property of guilt, then it might nonetheless be unjustifiable to seek it merely as retribution for an unfair emotion. Carlsson makes the following argument:

(a) to be blameworthy is to deserve to feel guilty,

(b) to feel guilty is to suffer, and

(c) one deserves to suffer for A only if A was under one's control.[16]

We can interpret (a) as the fittingness condition of anger. As such, it is also a necessary condition for the appropriateness of blame. Interpreted in this way, it would mean that for anger – and its action tendencies – to be appropriate responses to A is for the target to deserve to feel guilty about A. Premise b is highly plausible and needs no further interpretation. And if we accept (c), then it follows that anger is an appropriate response to A only if A was under the target's control. Which again runs into the no-control argument discussed in Section 4.1. In effect, whether this argument means the end for anger as an appropriate response to unfair emotions hangs on our acceptance of Premise c. Regrettably, Carlsson does not provide any further argument for this premise other than appealing to its intuitive plausibility.[17]

As I see it, we have two options here. Either, we reject Premise c on the basis of considerations such as the ones given by Bennett. Namely, we accept that the aim of retribution is a renewed commitment to a moral

good and to make the target feel guilty or remorseful, but not blanket suffering. Thereby, we can conclude that the target suffers in a morally meritorious way which is justified even in response to something uncontrolled like emotions. Or, if we find Premise c appealing enough to trump any theoretical considerations, then we have to accept that an episode of anger that aims at making its target feel guilt or any other painful emotion is never an appropriate response to a mere emotion.

Even if we grant that in the case of unfair emotions a desire for revenge is limited to a desire to pay the offender back *in kind* or to make them see the error of their ways, and that there might be moral benefits to this kind of practice, Nussbaum's more general critique of anger as a vengeful emotion still stands. However, in this case, the burden is shared with any theory of blame that casts anger as a central reactive attitude and the criticism does not fall on unfair emotions alone but is a criticism of responding with anger to actions as well. To try and avoid even this more general critique of anger, I turn to the second strategy promised above, namely, to deny that anger necessarily aims at revenge or retribution but rather is a way of protesting a wrong or aims at the target's recognition of one's rights and claims.

6.3.2 Protest and Recognition

There are a growing number of authors who disagree with the traditional portrayal of anger as primarily focused on revenge. Smith argues that blame is not a form of sanction or lust for revenge, but a way of expressing moral protest against wrongdoing.[18] While Smith argues that *blame* is a form of protest, something similar can be argued for *anger*. Namely, that anger is not an emotion that involves a desire for revenge, but rather the motivational tendency to protest wrongdoing. This is especially plausible if we accept anger as one of the reactive attitudes most central to the RA conception of blame.

Silva also draws into question the orthodox view of anger as involving a desire for revenge, and instead argues that anger is better conceived as pluralistic, and often involves a desire for recognition instead of revenge.[19] For one, she points to recent psychological research that shows that anger only aims at revenge in atypical cases, namely, when the subject feels that there is nothing to lose, and all other options are exhausted. In most typical cases, anger motivated the subject to confront the offender and seek to get their recognition of the wrong they have committed.[20] The desire for revenge typically only arises when all options to get any recognition seem unattainable. This interpretation bears some similarities to Bennett's, with the difference that it is recognition and not guilt that is sought – and thereby it might not necessarily involve suffering.

Does the motivational profile of anger make sense when construed as a form of protest or demand for recognition rather than a desire for revenge or retribution? Certainly, the confrontational part does. A protest does not work very well if the person at whom it is directed is never confronted with it. Hence, making the offender pay attention to the objections of the offended and recognize their claims is well justified. The aggressive or forceful nature of anger, on the other hand, is more worrisome. If it is interpreted as a desire for payback or violence, it is harder to defend it from her charges of magical thinking or status obsession.

However, there are less severe interpretations of the aggressive side of anger. Nussbaum's interpretation of anger makes anger seem more like hate, aiming to *proactively* harm or disadvantage another. When you hate someone, you don't require them to commit any moral offence to desire their misery. You might even want them to act morally objectionably, simply to give you an excuse to want them to feel bad. As such, hate seems to have an aim or desired outcome that is independent of whether the target already feels guilty or remorseful. In contrast to hate, protest- or recognition-accounts of anger conceive of anger as a *reactive* emotion, that responds to an already existing offence and is responsive to changes in the offender's attitudes, such as guilt or remorse.

Anger signals a readiness to use aggressive or forceful means to enforce adherence to a shared norm or value. As I have argued it in Sections 5.3 and 5.3.1, the offensive conduct represents a defection from basic human regard, the minimum expectation about how people are allowed to relate to others. Protesting such a defection signals to the offender that the subject will not tolerate it and that they are ready to try to enforce compliance. When getting angry, the subject does not only signal such a readiness, but actually feels an action readiness to confront the offender and forceful behavioural tendencies. After all, what better way is there to signal a readiness to use force than actually being motivated by the emotion to act forcefully?

Is your flatmate's anger an appropriate reaction to your joy? The first aspect of anger is its confrontational force. It would seem perfectly appropriate if your flatmate confronted you and demanded you answer for enjoying your neighbour's misery. You are answerable for your enjoyment because it reflects your underlying attitude of hate towards your neighbour and is not a random or inexplicable fluke. You did not take any happy-pills beforehand or have any condition that makes you enjoy random situations. Your joy is rationally connected to your other attitudes and therefore fulfils the answerability conditions.

Your flatmate is also not stepping out of line to hold you accountable on your neighbour's behalf, since what you display is a lack of basic human regard. A minimum regard for other people's well-being is something that we can justifiably expect any typical human agent to have, in virtue of a

shared basic relation as humans or moral agents. In effect, it is something we can expect people to value on third parties' behalf without overstepping our boundaries or unduly meddling in other people's relationships.

A second aspect of anger's motivational character is a forceful or aggressive tendency. Is aggression or a motivation to react with force and coercion appropriate? If it is in support of enforcing the requirements of basic human regard, it can be. The forceful tendencies of anger are often falsely identified as a desire for revenge. As mentioned, in many discussions of anger, it is portrayed as closely related to rage or even hatred, a desire to inflict pain or harm onto others in retribution for some slight or show of disrespect. I think this is a mischaracterization of anger that relies too much on an idea of blind rage, hatred, and ideas of vengeance. Rather, the motivation inherent in anger is closer to a desire to enforce adherence to a value or oppose defection from it. Anger does not directly aim at inflicting pain and suffering, but motivates the willingness to use force to prevent and oppose another person's malicious motives. In this capacity, it is not an instance of magical thinking, as Nussbaum[21] claims, that aggression can help address a wrong in the world. Neither does anger only make sense when seen as an expression of a vicious obsession with relative social status.[22] It rather reflects a concern for a moral good and a readiness to enforce other people's adherence to it.

To illustrate the difference between the more focused motivational profile of anger and that of a blindly aggressive emotion like hate or rage, it is useful to again consider how they relate to the relation between subject and target. As discussed in the last chapter, Section 5.3, Callard portrays anger as a reaction to a frustrated expectation within a co-valuing relationship, namely, the expectation that the target should share our values.[23] Anger occurs in a relation with someone that constitutively involves that both parties value certain goods, like each other's basic well-being or autonomy, which we could reasonably expect every moral human agent to share. If one party defects from valuing these goods while the other remains committed – for example by acting in violation of the other person's autonomy – this elicits a sense of frustration or dissonance in the adherent person between the relational expectation and the perceived defection. Anger is the felt reaction to this frustration, and hence arises out of a sense of powerlessness and urgency. Because we value a good like showing basic human regard for others, and expect to share this valuing with others, we are vulnerable to feel emotions like anger, hurt or disappointment when someone defects from this co-valuing.

This portrayal gets something important right about anger. Anger is not an emotion of blind hatred or bloodlust, that motivates us to inflict pain or harm on someone in any way possible. Rather, anger's more aggressive side is bound in service of addressing a perceived wrong. The felt readiness to

engage in forceful behaviour is simply a marshalling of our body's physical resources to defend an important good or counteract a threat to something we value. When directed at other people, this does not imply a desire to hurt them, but a willingness to hurt or coerce them when deemed necessary. This means that a great deal of situational awareness and discretion falls on the shoulders of the angry person. Anger is not as blunt as rage, it is much more complicated and complex than often granted.

We can see this in how targeted anger can be. In the above example, your flatmate, while she might be angry with you for enjoying your neighbour's misfortune, would probably not try to hurt or harm you as a form of payback. Rather, her aggressive behaviour will be limited to confronting you with your unacceptable hate for your neighbour. Forceful actions would mainly consist in holding you back when you try to avoid the confrontation and walk away, or shaking you by the shoulders when you try to change the subject or play down the importance of the situation. Of course, anger can get out of hand or spill over to affect uninvolved people. Your flatmate could hit the wall in frustration over your unwillingness to admit fault or balk at her friend later that day. But when something like that happens, her anger becomes just as irrational and potentially unfair as any other misplaced emotion. Appropriate anger in response to any offence will be focused and limited in the scope of its motivational tendencies.

In contrast to this, if your flatmate simply went on to hate you or feel rage towards you, she would stop caring about your attitude or whether you value other people's well-being. In such a case, she would possibly feel the same kind of joy at seeing you hurt as you did at your neighbour's misfortune in the above example. This is not so with anger. My suggestion is that the difference between hate and anger can be illustrated using Callard's framework in the following way: Hate reflects a defection from the underlying co-valuing relationship, while anger reflects the frustration and dissonance that comes from adhering to the co-valuing relationship while someone else defects from it. While both types of emotion can be responses to a wronging, hate and rage reflect a giving up the relationship, while anger reflects holding on to it. Applied to the above example, this would mean that if your flatmate is simply angry with you, she would feel motivated to confront you and aim at getting you to recognize your joy's unfairness. While if she simply came to hate you for it, she might no longer care for your well-being or that you genuinely come to see the error of your ways, but merely feel motivated to harm you or see you suffer.

In conclusion, emotions like hate or rage, which share some of the aggressive tendencies of anger, might very well be morally unjustified, especially in response to mere emotions. Anger, in contrast, does not share the same problematic aggressive tendencies. Rather, what aggressive behaviour it might motivate is focused on the perceived defection from an

important shared value, expresses protest at this defection, and the readiness to defend the value or oppose its violation. If this view of anger is right and anger is a fair and fitting response to a wronging, even in the form of an unfair emotion, then there is also no danger of a vicious circle in blaming others for genuinely unfair emotions. Resenting someone for an unfair emotion that reflects a disregard for you would dispose you to feel anger towards them, and that anger would count as a fair emotion, under those circumstances. Hence, resentment as the disposition to feel such anger is itself an appropriate attitude and does not count as an instance of disregard. Under these conditions, it would be fair to resent someone for an unfair emotion.

Notes

1 Deonna and Teroni, *The Emotions*, 2012.
2 Following the terminology of Deonna and Teroni, *The Emotions*, 8.
3 As I have discussed in Section 5.2.2.
4 For an argument against anger and indignation being the same species of emotion, see Drummond, "Anger and Indignation."
5 There are however arguments that this is merely an arational, see Hursthouse, "Virtue Theory and Abortion," 58; or even irrational, see Nussbaum, "Transitional Anger," 43–44, kind of response to inanimate objects; also Averill, "Studies on Anger and Aggression," 1149, reports a tendency of the subjects to personalise inanimate objects when they are the target of anger. In that case, it might be open to question whether anger towards inanimate objects should count as evidence of non-moralised anger.
6 Strawson, "Freedom and Resentment," 1962, section 5.
7 Specifically the version put forward by Strawson, "Freedom and Resentment," 1993, which I have adapted.
8 See Nussbaum, "Transitional Anger"; also Srinivasan, "The Aptness of Anger"; Pettigrove, "Meekness and 'Moral' Anger."
9 Nussbaum, "Transitional Anger," 47–48.
10 Nussbaum, "Transitional Anger," 49.
11 See Smith, "Moral Blame and Moral Protest"; Hieronymi, "The Force and Fairness of Blame"; Silva, "Anger and Its Desires"; Silva, "Is Anger a Hostile Emotion?"
12 See Nussbaum, "Transitional Anger," as discussed above in section 6.3.
13 Ï will use the terms *revenge* and *retribution* roughly interchangeably here.
14 Bennett, "The Varieties of Retributive Experience," 150.
15 This might seem closely connected to the doctrine of double-effect popularized by Aquinas. Despite there being a large debate on the validity of such distinctions, I will not go on to discuss this here, and simply assume that the distinction is meaningful in this case.
16 Carlsson, "Blameworthiness as Deserved Guilt," 89.
17 Carlsson, "Blameworthiness as Deserved Guilt," 91.
18 Smith, "Moral Blame and Moral Protest."
19 Silva, "Anger and Its Desires"; Silva, "Is Anger a Hostile Emotion?"
20 This is a point raised in feminist philosophy on the value of anger. The characterization of anger as inherently violent often makes it easier to dismiss rather

than recognizing it as a powerful tool of protest and resistance. See, e.g. Frye, "A Note on Anger"; Lorde, "The Uses of Anger."

21 As argued above, and see Nussbaum, "Transitional Anger," 47–48.

22 See Nussbaum, "Transitional Anger," 49.

23 Callard, "The Reason to Be Angry Forever," 126–27 and 130–31.

7 Conclusion

Throughout the previous chapters, I have discussed unfair instances of emotions, what makes them unfair, and under what conditions we are to blame for them. At the beginning of the book, I introduced the phenomenon that we seem to treat emotions like they are proper objects of moral evaluation. Moreover, we can distinguish unfair instances of emotions that wrong the people towards whom they are directed from emotions that are simply overall morally problematic. I argued that we should take this everyday phenomenon seriously and not discount unfairness criticism of emotions as irrational or morally misguided. Instead, we should try to give a theoretical account for them that fits into a broader framework of moral philosophy, which was the aim of this book.

I have argued that the criticism of an emotion as unfair is distinct from other types of criticism, such as unfittingness, inconsistency, imprudence, or general immorality. In contrast to these other types of criticism, an emotion is unfair if its action tendencies constitute a morally objectionable hazard directed at its target that is not intrinsically morally justified by the emotion's fit-making conditions. This entails, for one, that an emotion's inherent action tendencies pose a threat to the interests or well-being of the target of the emotion. Furthermore, such a hazardous emotion is unfair only if it is also unfitting. This is because unfairness relies on the inherent justification of the emotion, which is given by its fit-making properties. Hence, if an emotion's fit-making conditions are met and at the same time provide a *pro tanto* moral justification for the emotion's inherent action tendencies, the emotion would not be unfair. In the absence of this justifying factor, such a hazardous emotion is unfair to its target.

I have further argued that it is unlikely that we have the relevant voluntary control over our emotions to justify any sanctions on the basis of unfair emotions. However, we do have a rational type of control over our emotions, in the sense that they are responsive to reasons and reflect our underlying cares and values. This type of rational control is sufficient to justify holding subjects answerable for their emotions. I have made the

DOI: 10.4324/9781003498032-7

case that rational control is also the type of control required for the way of holding people accountable proposed by reactive attitudes (RA) theory. To hold someone accountable for an unfair emotion means being disposed to respond to it with reactive emotions, such as anger or indignation. On this basis, I have developed an account of what it entails to blame someone for an unfair emotion. I argue that blaming someone for an unfair emotion means resenting them for it. And resenting someone for an unfair emotion means being disposed to feel a range of reactive emotions – namely anger or indignation – towards the target of blame, on the basis of the unfair emotion reflecting a violation of the expectation of basic human regard towards either you or someone else.

With these two lines of argument, I have addressed the two large problems for a theoretical account of unfair emotions, which I have introduced at the very beginning of the book. With that, I have shown that the sceptical worry that our everyday practice of treating emotions as unfair is either irrational or morally misguided is not as pressing as it might initially seem. Hence, we don't need to discount this practice or try to explain it away. Instead, it is possible to account for it and integrate it into our broader ethical theories.

7.1 Outlook

There are several outstanding issues that I have only touched on in this book. One of them concerns the importance of uncertainty in questions about the unfairness of emotions and our reactive attitudes towards them. In Section 6.2, I have noted some of the epistemic difficulties we can be faced with when considering whether anger or resentment are justified responses to take towards someone. I already mentioned that resentment can be *misplaced*, when it is based on false information. But, of course, even if an instance of resentment is misplaced, it is often hard for the resenter to realize this. In most cases, there are many uncertainties about the underlying attitudes of the person feeling an unfair emotion, or even in judging whether they in fact do feel the perceived emotion.

While this book focused on the question of what a standard of unfairness in regard to emotions looks like, and whether it could ever be appropriate to hold people accountable for them, it did not examine the epistemic standards of what kind of evidence we would need to be justified in ascribing unfair emotions to others, or to hold them accountable for them. However, we can split this issue of uncertainty into two parts.

First, there are similar uncertainties that we would also have with responding to wrongful actions. These include questions about the motivation and underlying attitudes of the perpetrator. Do they disregard my well-being? Do they even despise me? These kinds of questions have to be

addressed by any theory of justified blame, and are not specific to blaming emotions.

Second, it is more difficult to know whether someone actually does experience an unfair emotion, than whether they have performed a wrongful action. Actions are typically much more clearly observable than emotions. This is not always the case, since actions can also be obscured and emotions sometimes cannot be hidden from observers who are familiar with the person feeling them. It is rather a matter of degree that it is harder to know what someone feels than to know what they have done. In either case, whether reacting with anger is an appropriate response or not boils down to what counts as a sufficient degree of certainty and to how we understand anger.

In Section 6.3.2, I have argued, that anger is not always and not necessarily the violent, destructive emotion it has come to be viewed as. Rather, anger can be seen as a kind of protest, and when expressed signal a demand for recognition and respect. In this, still confrontational but less destructive, view of anger, it can be a justified response to a perceived unfair emotion in someone else. The difference between an account of anger that understands it as a form of protest, and an account that understands it as aiming to punish, is like the difference between an accusation and a final judgement.

The protest approach leaves room for error and uncertainty. If anger is necessarily violent and retributive, there seems to be a much higher necessity to be certain that a wrong has occurred, or in our case, that an unfair emotion is felt. Otherwise, one runs a high risk of being motivated to punish someone for nothing. However, if anger is understood as a type of protest, then it only motivates confrontation and addressing the person who supposedly feels an unfair emotion. The motivation to confront is arguably much less epistemically demanding than the motivation to punish, since it inherently aims at engagement. If you are confronted about an unfair emotion you did not really feel, and you defend yourself, clarifying the falsity of the accusation, this might completely resolve the issue. However, if you came to know that the same person intended to punish you for the same mistaken interpretation of your emotional state, you might justifiably feel pre-judged.

This being said, the topic of epistemic justification of the reactive attitudes, like anger, is a broad one, and cannot be discussed to the degree it deserves in the final chapter of this book, but it is also not the central topic at issue in this book. What has been addressed are the conditions under which reactive attitudes would be appropriate responses to unfair emotions, given that one has a sufficient degree of certainty about the underlying attitudes that gave rise to them. When and under what conditions such a degree is reached is another matter that requires further discussion.

A second outstanding issue is the discussion of the particular reactive attitudes that are involved in holding people accountable for unfair emotions. Discussions such as the one on the aims of anger, in Section 6.3, show the importance of how we understand our responses to wronging. In this book, I have taken the position of Silva and others, that anger is not, or at least not only or primarily, a retributive emotion that aims at revenge or punishment. But rather that it is an emotion of protest and confrontation that aims at recognition and admission of fault. However, this view can be and has been challenged. Authors like Nussbaum[1] have argued that this way of portraying anger amounts to a sanitization of what is actually a violent, retributive emotion.

As a point of unity, both authors are in agreement that the issue of how to best characterize an emotion like anger cannot be done by philosophy alone and requires psychological study. While I have here relied on the work done by Silva[2] to provide a philosophically convincing case based on a thorough psychological background, this is still an area of ongoing debate and research. Therefore, it is probably the case that anger is more complex or more difficult to characterize than suggested here. Hence, there is always room to argue that anger in response to a perceived unfair emotion, or even in response to anything, may never be justified. This, as well, is a vastly larger issue, that deserves further and ongoing attention, especially from an empirically informed perspective that aims at doing normative philosophy of emotion.

A third issue this book does not attempt to address, but lies at the heart of unfairness of emotions, is to provide an account of the fittingness of emotion. In Section 3.2, I have argued that the unfittingness of an emotion is a necessary condition for its unfairness. While I have given a general explanation of what is commonly meant by *unfittingness* in Section 2.1, I did not take a substantive position on what fittingness is or how the fittingness conditions of any given emotion are constituted. For one, this is an intentional decision in order to offer an account of unfairness that should be compatible with many different, if not most, notions of fittingness. For another, the question of what fittingness is and how it is constituted is itself a large debate that runs somewhat orthogonal to the question of unfairness.

However, there might be some questions on unfairness that are centrally affected by the account of fittingness one holds. On the one hand, whether we hold that there is an objective, stance-independent standard of fittingness for any given emotion would also imply that we should think of unfairness as involving such a stance-independent standard. On the other hand, if we take a stance-dependent view of fittingness, this would imply stance-dependence for unfairness as well. Which standard we assume will influence how we understand and approach disagreements about the fairness of an emotion. Under a stance-independent interpretation, we might

find that disagreements about whether an instance of anger is unfair to be decidable on one side or the other, similar to disagreements about whether the film *Alien* was released in 1978 or 1979. While if we take a stance-dependent view of fittingness, disagreements about the fairness of emotions might be more akin to disagreements whether vanilla or chocolate ice cream tastes better.

While these theoretical commitments have practical implications about how to understand and address disagreements about the fairness of an emotion, they are not specific to emotions. Rather, these are widely debated meta-ethical questions that have to be addressed for any given normative standard, be it fittingness, fairness, duty, wrongness, value, or even rationality itself. In effect, similar issues occur for questions of the rightness of actions just as much as for the fairness of emotion. But these issues are of a much wider concern than the topic of this book and while they are relevant and important in and of themselves, I cannot do them justice in these last pages of the concluding chapter. All I want to stress is that it is worthwhile pursuing them further, in particular with respect to the fittingness and fairness of emotions.

7.2 Upshots

The main positive upshot of this book is that it provides a theoretical account for when emotions are unfair to their targets, when we are to blame for such an unfair emotion, and when it is a blameless instance of an unfair emotion.

One benefit of this account is that it provides a guide and criteria for discussing when an emotion is unfair and when the subject can be blamed for it. This also provides criteria for when we are not to blame for an emotion and when such criticism would itself be a morally objectionable infringement into our private mental space. For example, when you are disgusted by an ice-cream, this would not count as unfair since it does not pose a directed moral hazard to its target. In addition, nobody should blame you for this emotion since it does not – at least in most plausible interpretations – reflect an attitude of disregard towards its target. Hence, it would be inappropriate to criticize you for being disgusted by ice cream, and neither should such criticism take on an angry or resentful dimension. In contrast, disgust of other people might be unfair and sometimes even a valid basis for blame. For example, if your disgust stems from and reflects a disregard for the humanity or well-being of those people. In such a case, not only are you criticizable as being unfair to others, but also a fair target of resentment on their behalf.

An account of unfair emotion can also give a more nuanced response to the apparent issues that come with morally judging someone's emotions.

A common response to people feeling guilt or shame at their own anger under conditions of oppression, as brought up by authors like Lorde[3] or Silva,[4] is to argue that we should never feel guilt or shame for our emotions and neither blame nor criticize others for theirs. But such emotions of shame and guilt over how we feel sometimes do arise, and to brush them away as irrational invites us to reflect less on the reasons we might have for them. But it is exactly in these situations that we should examine the reasons for why we feel this way, to resolve the issue that gave rise to them in the first place. With a notion of unfair emotion at hand, we can better start addressing these issues. Not on the basis that we should never feel guilt or shame for our own emotions, because sometimes we should, but on the basis of the moral injustice that justifies the anger.

The account presented here is also well-placed to respond to the charges of thought-policing that the very mention of morally judging emotions might give rise to. The delineation between when our emotions are open to moral criticism by others, and when they are none of their business, seems to capture a good balance between the social and the private side of emotions. On the one hand, our emotions are features of our relationships to others and relevantly structure and shape those. On the other hand, our emotions are very intimate parts of our mental lives, and external restrictions or chastisement of them can quickly become invasive or even abusive. There might still be other types of moral criticism that can be made, even against fair emotions. However, in such cases the criticism can always be accompanied by an acknowledgement that while the emotion is not all–things–considered morally optimal, there is a kernel of intrinsic moral justification – or fairness – to it.

Another benefit of my account of unfair emotions is its usefulness for philosophical debates. One debate that can potentially benefit from such an account, which I have already mentioned in the introduction, is the debate around the reactive attitudes theory of moral responsibility. The potential benefit I mentioned in Section 1.4 is that a standard of fairness cannot only dispel the weaker worry that reactive emotions are unfitting responses to wrongdoing – meaning that they are misrepresenting or falsely responding to the circumstances – but the stronger worry that it is morally inappropriate to get angry with or resent someone for their conduct. My account of unfair emotions can answer this stronger worry by showing that emotions like anger or resentment can be at least intrinsically morally justified in response to conduct that reflects the agent's disregard for others. It therefore provides a standard of moral justification on top of considerations of the fittingness of these reactive emotions. How this can be integrated into a wider theory of moral responsibility is the work of further investigation.

I hope to have shown that if you find someone else's emotion towards you unfair, then you might not be completely irrational or morally misguided, but that there are good arguments to support this appearance. At the same time, our criticism of other people's emotional lives has to remain measured and sensitive to context. We are only really worthy of resentment for such an unfair emotion if it reflects deeper moral shortcomings of our attitudes towards others. Therefore, if I get angry with you out of a vicious motivation, it would be fair for you to resent me for it.

Notes

1 Nussbaum, *Anger and Forgiveness*.
2 Silva, "Anger and Its Desires"; Silva, "Is Anger a Hostile Emotion?"
3 Lorde, "The Uses of Anger."
4 Silva, "Anger and Its Desires."

Bibliography

Adams, Robert Merrihew. *A Theory of Virtue: Excellence in Being for the Good*. Oxford University Press, 2006.

Averill, James R. "Studies on Anger and Aggression: Implications for Theories of Emotion." *American Psychologist* 38, no. 11 (1983): 1145–60. https://doi.org/bw9r89.

Barrett, Lisa Feldman. *How Emotions Are Made: The Secret Life of the Brain*. Houghton Mifflin Harcourt, 2017.

Bell, Macalester. "Globalist Attitudes and the Fittingness Objection." *The Philosophical Quarterly* 61, no. 244 (July 2011): 449–72. https://doi.org/c6c7z3.

Bell, Macalester. *Hard Feelings: The Moral Psychology of Contempt*. Oxford: Oxford University Press, 2013.

Bennett, Christopher. "The Varieties of Retributive Experience." *The Philosophical Quarterly* 52, no. 207 (2002): 145–63. https://doi.org/b72rrf.

Bittner, Rüdiger. "Is It Reasonable to Regret Things One Did?" *The Journal of Philosophy* 89, no. 5 (May 1992): 262–73. https://doi.org/dpt9px.

Brady, Michael S. "The Irrationality of Recalcitrant Emotions." *Philosophical Studies* 145, no. 3 (September 2009): 413–30. https://doi.org/c56sxn.

Broome, John. "Wide or Narrow Scope?" *Mind* 116, no. 462 (April 2007): 359–70. https://doi.org/dd8k4q.

Callard, Agnes. "The Reason to Be Angry Forever." In *The Moral Psychology of Anger*, edited by M. Cherry and O. Flanagan, 123–37. London: Rowman & Littlefield International, 2017.

Carlsson, Andreas Brekke. "Blameworthiness as Deserved Guilt." *The Journal of Ethics* 21, no. 1 (March 2017): 89–115. https://doi.org/gfkzmf.

Cornell, Nicolas. "Wrongs, Rights, and Third Parties." *Philosophy & Public Affairs* 43, no. 2 (March 2015): 109–43. https://doi.org/gfkzn4.

Cox, Damian, and Michael Levine. "Believing Badly." *Philosophical Papers* 33, no. 3 (November 2004): 309–28. https://doi.org/dtkfgx.

D'Arms, Justin. "Value and the Regulation of the Sentiments." *Philosophical Studies* 163, no. 1 (March 2013): 3–13. https://doi.org/gf23xg.

D'Arms, Justin, and Daniel Jacobson. "Sentiment and Value." *Ethics* 110, no. 4 (July 2000): 722–48. https://doi.org/10.1086/233371.

D'Arms, Justin, and Daniel Jacobson. "The Moralistic Fallacy: On the 'Appropriateness' of Emotions." *Philosophy and Phenomenological Research* 61, no. 1 (2000): 65–90. https://doi.org/fpbd3f.

D'Arms, Justin, and Daniel Jacobson. "VIII. The Significance of Recalcitrant Emotion (or, Anti-Quasijudgmentalism)." *Royal Institute of Philosophy Supplements* 52 (March 2003): 127–45. https://doi.org/df46vf.

D'Arms, Justin, and Daniel Jacobson. "Wrong Kinds of Reason and the Opacity of Normative Force." In *Oxford Studies in Metaethics*. Vol. 9, edited by Russ Shafer-Landau, 215–44. Oxford University Press, 2014. https://doi.org/10.1093/acprof:oso/9780198709299.003.0009.

Darwall, Stephen L. "Two Kinds of Respect." *Ethics* 88, no. 1 (October 1977): 36–49. https://doi.org/ctm6mk.

Darwall, Stephen L. "Respect and the Second-Person Standpoint." *Proceedings and Addresses of the American Philosophical Association* 78, no. 2 (2004): 43–59. https://doi.org/d88hjw.

Darwall, Stephen L. *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge, MA: Harvard University Press, 2006.

de Sousa, Ronald. *The Rationality of Emotion*. Cambridge, MA: MIT Press, 1987.

Dennett, Daniel C. *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge, MA: MIT Press, 2015.

Deonna, Julien A., and Fabrice Teroni. "From Justified Emotions to Justified Evaluative Judgements." *Dialogue* 51, no. 01 (2012): 55–77. https://doi.org/gfrjbk.

Deonna, Julien A., and Fabrice Teroni. *The Emotions: A Philosophical Introduction*. London; New York: Routledge, 2012.

Deonna, Julien A., and Fabrice Teroni. "Emotions as Attitudes." *Dialectica* 69, no. 3 (September 2015): 293–311. https://doi.org/gfkznm.

Deonna, Julien A., and Fabrice Teroni. "Which Attitudes for the Fitting Attitude Analysis of Value?" *Theoria* 87 (June 2021): 1099–122. https://doi.org/gk85dp.

Derek, Parfit. *Reasons and Persons*. Oxford: Oxford University Press, 1984.

Döring, Sabine A. "Seeing What to Do: Affective Perception and Rational Motivation." *Dialectica* 61, no. 3 (2007): 363–94. https://doi.org/ff8c8p.

Döring, Sabine A. "Why Recalcitrant Emotions Are Not Irrational." In *Emotion and Value*, edited by Sabine Roeser and Cain Todd, 124–36. Oxford University Press, 2014. https://doi.org/10.1093/acprof:oso/9780199686094.003.0008.

Döring, Sabine A. "What's Wrong with Recalcitrant Emotions? From Irrationality to Challenge of Agential Identity." *Dialectica* 69, no. 3 (2015): 381–402. https://doi.org/gfkznb.

Drummond, John J. "Anger and Indignation." In *Emotional Experiences: Ethical and Social Significance*, edited by John J. Drummond and Sonja Rinofner-Kreidl. London; New York: Rowman & Littlefield, 2017.

Echeverri, Santiago. "Emotional Justification." *Philosophy and Phenomenological Research* 98, no. 3 (2019): 541–66. https://doi.org/gm5qt7.

Feinberg, Joel, and Jan Narveson. "The Nature and Value of Rights." *The Journal of Value Inquiry* 4, no. 4 (1970): 243–60. https://doi.org/fvwxqt.

Fischer, John Martin. *My Way: Essays on Moral Responsibility*. Oxford University Press, 2006.

Fischer, John Martin, and Mark Ravizza. *Responsibility and Control*. Cambridge: Cambridge University Press, 1998.

Frankfurt, Harry G. "Alternate Possibilities and Moral Responsibility." *The Journal of Philosophy* 66, no. 23 (1969): 829–39. https://doi.org/d92vmc.

Fricker, Miranda. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press, 2007.

Fricker, Miranda. "What's the Point of Blame? A Paradigm Based Explanation." *Noûs* 50, no. 1 (2016): 165–83. https://doi.org/gf2r8d.

Frijda, Nico H. *The Emotions*. Cambridge: Cambridge University Press, 1986.
Frye, Marilyn. "A Note on Anger." In *The Politics of Reality*, 176. Crossing Press, 1983.
Gertken, Jan, and Benjamin Kiesewetter. "The Right and the Wrong Kind of Reasons." *Philosophy Compass* 12, no. 5 (2017): e12412. https://doi.org/ghh5tz.
Graham, Peter A. "A Sketch of a Theory of Moral Blameworthiness." *Philosophy and Phenomenological Research* 88, no. 2 (2014): 388–409. https://doi.org/gfvnc3.
Gross, James J. "The Emerging Field of Emotion Regulation: An Integrative Review." *Review of General Psychology* 2, no. 3 (1998): 271–99. https://doi.org/bg3k9k.
Gross, James J. "Emotion Regulation: Affective, Cognitive, and Social Consequences." *Psychophysiology* 39, no. 3 (2002): 281–91. https://doi.org/dhr5xt.
Gross, James J. *Handbook of Emotion Regulation*. Edited by James J. Gross. 2nd ed. New York: Guilford Press, 2014.
Hart, H. L. A. "Legal Responsibility and Excuses." In *Punishment and Responsibility: Essays in the Philosophy of Law*, 28–53. Oxford University Press, 2008.
Hart, H. L. A. *Punishment and Responsibility: Essays in the Philosophy of Law*. Oxford University Press, 2008.
Hieronymi, Pamela. "Articulating an Uncompromising Forgiveness." *Philosophy and Phenomenological Research* 62, no. 3 (May 2001): 529–55. https://doi.org/chtjwr.
Hieronymi, Pamela. "The Force and Fairness of Blame." *Philosophical Perspectives* 18, no. 1 (December 2004): 115–48. https://doi.org/c5vfqv.
Hieronymi, Pamela. "The Wrong Kind of Reason." *The Journal of Philosophy* 102, no. 9 (2005): 437–57. https://doi.org/f3f92h.
Hieronymi, Pamela. "Controlling Attitudes." *Pacific Philosophical Quarterly* 87, no. 1 (2006): 45–74. https://doi.org/bdwxxf.
Hieronymi, Pamela. "Responsibility for Believing." *Synthese* 161 (2008): 357–73. https://doi.org/b54mq3.
Hills, Alison. "Moral Testimony and Moral Epistemology." *Ethics* 120, no. 1 (October 2009): 94–127. https://doi.org/c3366g.
Hohfeld, Wesley Newcomb. "Some Fundamental Legal Conceptions as Applied in Judicial Reasoning." *The Yale Law Journal* 23, no. 1 (1913): 16–59. https://doi.org/fpz2td.
Hume, David. *A Treatise of Human Nature: A Critical Edition*. Edited by David Norton and Mary Norton. Oxford: Clarendon Press, 2007. Reprint, 1738.
Hurka, Thomas. *Virtue, Vice, and Value*. Oxford University Press, 2001.
Hursthouse, Rosalind. "Virtue Theory and Abortion." *Philosophy & Public Affairs* 20, no. 3 (July 1991): 223–46.
James, William. "What Is an Emotion?" *Mind* 9, no. 34 (1884): 188–205. https://doi.org/ch377c.
Kant, Immanuel. "Grundlegung Zur Metaphysik Der Sitten." In *Kants Werke*, Akademie Textausgabe. Vol. 4, 385–464. Berlin: Walter de Gruyter & Co., 1968.
Korsgaard, Christine M. "Kant's Formula of Humanity." *Kant-Studien* 77, no. 1–4 (January 1986): 183–202. https://doi.org/fg8kdj.
Lorde, Audre. "The Uses of Anger." *Women's Studies Quarterly* 25, no. 1/2 (1997): 278–85. www.jstor.org/stable/40005441.
Macnamara, Coleen. "Holding Others Responsible." *Philosophical Studies* 152, no. 1 (January 2011): 81–102. https://doi.org/c8m27m.

Maguire, Barry. "There Are No Reasons for Affective Attitudes." *Mind* 127, no. 507 (July 2018): 779–805. https://doi.org/gfkznv.
McGeer, Victoria. "Scaffolding Agency: A Proleptic Account of the Reactive Attitudes." *European Journal of Philosophy* 27, no. 2 (2019): 301–23. https://doi.org/ggxrdv.
McHugh, Conor, and Jonathan Way. "Fittingness First." *Ethics* 126, no. 3 (2016): 575–606. https://doi.org/gfkzn5.
McKenna, Michael. *Conversation and Responsibility*. Oxford, New York: Oxford University Press, 2012.
Menges, Leonhard. *Moralische Vorwürfe*. Berlin, Germany: De Gruyter, 2017.
Na'aman, Oded. "The Fitting Resolution of Anger." *Philosophical Studies* (June 2019). https://doi.org/ggk3b8.
Na'aman, Oded. "The Rationality of Emotional Change: Toward a Process View." *Noûs* 55 (2019): 245–69. https://doi.org/ggkd4s.
Nagel, Thomas. "Death." *Noûs* 4, no. 1 (February 1970): 73. https://doi.org/fpt6sk.
Nozick, Robert. "On the Randian Argument." *The Personalist* 52, no. 2 (1971): 282–304. https://doi.org/10.1111/j.1468-0114.1971.tb08926.x.
Nozick, Robert. *Anarchy, State, and Utopia*. New York: Basic Books, 1974.
Nussbaum, Martha C. "Emotions as Judgments of Value and Importance." In *Thinking About Feeling: Contemporary Philosophers on Emotions*, edited by Robert C. Solomon, 183–99. New York: Oxford University Press, 2004.
Nussbaum, Martha C. "Transitional Anger." *Journal of the American Philosophical Association* 1, no. 1 (2015): 41–56. https://doi.org/gg7mkq.
Nussbaum, Martha C. *Anger and Forgiveness: Resentment, Generosity, Justice*. Oxford: Oxford University Press, 2016.
Owens, David J. *Shaping the Normative Landscape*. Oxford: Oxford University Press, 2012.
Pettigrove, Glen. "Meekness and 'Moral' Anger." *Ethics* 122, no. 2 (January 2012): 341–70. https://doi.org/gmxbcq.
Prinz, Jesse. *Gut Reactions: A Perceptual Theory of Emotion*. Oxford: Oxford University Press, 2004.
Rabinowicz, Wlodek, and Toni Rønnow-Rasmussen. "The Strike of the Demon: On Fitting Pro-Attitudes and Value." *Ethics* 114, no. 3 (April 2004): 391–423. https://doi.org/bgsxq5.
Scanlon, Thomas M. *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge, MA: Harvard University Press, 2008.
Scanlon, Thomas M. "Giving Desert Its Due." *Philosophical Explorations* 16, no. 2 (June 2013): 101–16. https://doi.org/gf6bqc.
Scarantino, Andrea. "The Motivational Theory of Emotions." In *Moral Psychology and Human Agency*, edited by Justin D'Arms and Daniel Jacobson, 156–85. Oxford University Press, 2014.
Scarantino, Andrea. "Do Emotions Cause Actions, and If So How?" *Emotion Review* 9, no. 4 (October 2017): 326–34. https://doi.org/gfkznt.
Shoemaker, David. "McKenna's Quality of Will." *Criminal Law and Philosophy* 9, no. 4 (December 2015): 695–708. https://doi.org/gfxsdj.
Silva, Laura. "The Rationality of Anger." Doctoral Thesis. UCL (University College London), 2019.
Silva, Laura. "Anger and Its Desires." *European Journal of Philosophy* 29, no. 4 (2021). https://doi.org/gmj5bd.

Silva, Laura. "Is Anger a Hostile Emotion?" *Review of Philosophy and Psychology* (May 2021). https://doi.org/gjx7dz.
Silva, Laura. "The Epistemic Role of Outlaw Emotions." *Ergo* (2021): 39. https://doi.org/gp6kz9.
Singer, Peter. "Famine, Affluence, and Morality." *Philosophy and Public Affairs* 1, no. 3 (1972): 229–43.
Sliwa, Paulina. "In Defense of Moral Testimony." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 158, no. 2 (2012): 175–95. https://doi.org/gg5vr7.
Smart, J. J. C. "Free-Will, Praise and Blame." *Mind* 70, no. 279 (May 1961): 291–306. https://doi.org/fmjxfk.
Smith, Angela M. "Responsibility for Attitudes: Activity and Passivity in Mental Life." *Ethics* 115, no. 2 (January 2005): 236–71. https://doi.org/cb6p27.
Smith, Angela M. "Attributability, Answerability, and Accountability: In Defense of a Unified Account." *Ethics* 122, no. 3 (2012): 575–89. https://doi.org/gfkzm3.
Smith, Angela M. "Moral Blame and Moral Protest." In *Blame: Its Nature and Norms*, edited by D. Justin Coates and Neal A. Tognazzini, 27–48. Oxford: Oxford University Press, 2013.
Solomon, Robert C. "On Emotions as Judgments." *American Philosophical Quarterly* 25, no. 2 (1988): 183–91. www.jstor.org/stable/20014237.
Srinivasan, Amia. "The Aptness of Anger." *Journal of Political Philosophy* 26, no. 2 (2018): 123–44. https://doi.org/gfkzns.
Strawson, Peter F. "Freedom and Resentment." *Proceedings of the British Academy* 48 (1962): 1–25.
Strawson, Peter F. "Freedom and Resentment." In *Perspectives on Moral Responsibility*, edited by John Martin Fischer and Mark Ravizza, 45–66. New York: Cornell University Press, 1993.
Talbert, Matthew. "Moral Responsibility." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University, Winter 2019.
Tamir, Maya, Oliver P. John, Sanjay Srivastava, and James J. Gross. "Implicit Theories of Emotion: Affective and Social Outcomes Across a Major Life Transition." *Journal of Personality and Social Psychology* 92, no. 4 (2007): 731–44. https://doi.org/cnmxn3.
Tappolet, Christine. *Emotions, Values, and Agency*. Oxford: Oxford University Press, 2016.
Thompson, Michael. "What Is It to Wrong Someone? A Puzzle About Justice." In *Reason and Value*, edited by R. Jay Wallace, Philip Pettit, Samuel Scheffler, and Michael Smith, 333–84. Oxford: Oxford University Press, 2004.
Thomson, Judith Jarvis. *The Realm of Rights*. Cambridge, MA: Harvard University Press, 1990.
Tognazzini, Neal, and D. Justin Coates. "Blame." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University, Summer 2021.
Wallace, R. Jay. *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press, 1994.
Wallace, R. Jay. "Dispassionate Opprobrium: On Blame and the Reactive Sentiments." In *Reasons and Recognition: Essays on the Philosophy of T.M. Scanlon*, edited by R. Jay Wallace, Rahul Kumar, and Samuel Freeman, 384–85. Oxford: Oxford University Press, 2011.

Wallace, R. Jay. *The Moral Nexus*. Princeton University Press, 2019.
Wallace, R. Jay. "Trust, Anger, Resentment, Forgiveness: On Blame and Its Reasons." *European Journal of Philosophy* 27, no. 3 (2019): 537–51. https://doi.org/ggkd4t.
Watson, Gary. "Two Faces of Responsibility." *Philosophical Topics* 24, no. 2 (1996): 227–48. https://doi.org/fz8btv.
Wolf, Susan. "Blame, Italian Style." In *Reasons and Recognition: Essays on the Philosophy of T.M. Scanlon*, edited by R. Jay Wallace, Rahul Kumar, and Samuel Freeman, 332–47. Oxford University Press, 2011. https://doi.org/10.1093/acprof:oso/9780199753673.003.0014.

Index

Note: Page numbers in *italic* indicate a figure on the corresponding page.

accountability: answerability and 88–93; attribution and 88–90; blame and 87–97, 103; co-valuational expectations and 108–15; to expectations, holding others 101–4; fitting reactive emotions and 113–15; holding people accountable and 100, 138; human regard and, basic 111–13; methods of 87–97; moral 101; moral obligation and 105–6; negative 100–101; overview 100; quality of will and 106–8; reactive attitude theory and 100–101; reactive emotions and 103–8; resentment and 96–7, 100–101, 118; social sanctions and 88–9, 93–6; for unfair emotions 140; wrongs/wronging and 70–2; *see also* control; responsibility
action readiness 54
action tendencies 66–7
amusement 40, 53, 57–9, 61, 63, 114
anger: aim of 126–35; complexity of 140; critique of 41–2; expression of 11; fair 114–15, 122–3, 126–8, 134–5, 139; indignation versus 122–3; indirect 11; misplaced 37, 61, 72n12; as moral hazard 50–2; motivational profile of 126, 132–3; Nussbaum's interpretation of 132; offence and, response to 126; portrayal of, typical 117; protest and, moral 131–5, 139–40; punishment and 77; recognition and 131–5; regret about 41; resentment versus 118–23; responsibility for 92; retribution and 128–31; revenge and 50–1, 117, 128–31; targeted 134; as violent emotion 118; warranted 38
answerability 88–93
attitudes: evaluative 112–13; non-moral 104–8; non-moral reactive 104–8; non-reactive 104; personal reactive 110; reactive 88–9, 96–7, 101, 104–8, 139–40; theory about 24
attitudinal theory 24
attribution 88–90

bad-being of a person 103
basic human regard 111–13
behavior, emotional 3, 53–7
beliefs 50–1, 84–5
Bell, Macalester 38–40
Bennett, Christopher 129, 131
Bittner, Rüdiger 34
blamability/blame: accountability and 87–97, 103; aim of 94; answerability and 88–93; attribution and 88–90; control and 74–5; definition of 87; for emotions and 74; features of 87–8; implication of 88; negative reaction of 87; protest and, moral 131; reactive attitudes and 88–9, 96–7; reactive emotions and 114; resentment and 75, 99n52, 101, 126; responsibility and 90; social sanctions and 88–9, 93–6

bodily feelings 2, 54
Brady, Michael S. 32

Callard, Agnes 86, 110, 134
Carlsson, Andreas Brekke 94
claims and wronging 68–70
comparative fairness 45, 47
comparative unfairness 45–6
conditionality feature of unfair emotions 48, 57–63
contempt 38–9, 104
control: accountability and 77–80; blame and 74–5; direct 80–2, 84; of emotions 80–7, 137–8; evaluative 84; extrinsic reasons for 91; indirect 80, 83–4; intrinsic reasons for 91; managerial 84; no problem of 75–7; rational 81, 84–5; requirements, disentangling 77–80, *78*, *79*; resentment and 97; responsibility and 78, *78*, 97n4; voluntary 75, 80–4, 94
Cornell, Nicolas 70–1
co-valuational expectations 108–15
Cox, Damian 50
criticisms of emotions: inconsistency 22, 30–3; moral 4–5, 22, 37–42; prudential 33–6; *see also* questions applied to criticisms of emotions; unfairness criticism of emotions; unfittingness criticism of emotions

D'Arms, Justin 57, 83
debunking strategy 6, 9
Deonna, Julien A. 65, 118
de Sousa, Ronald 66
directed wrongs 67–72, 74–5
direct voluntary control 80–2, 84
disappointment 60
disgust 141
dispositions, emotional 103, 118–21, 135
dis-valuing the bad 39
duty 69, 71, 73n19

Echeverri, Santiago 25–6
emotional dispositions 103, 118–21, 135
emotional episodes 118–21
emotions: attitudinal theory of 24; behavior/motivation associated with 3, 53–7; blame for 74; bodily feelings and 2, 54; cake scenario and 7–8; control over 80–7, 137–8; definition of 2–4; directed 65, 67–8; directedness of 63; as evaluative mental states 3; expression of 8–10; fair 5, 114–15; features of 2–4; fitting reactive 113–15; focus on object/event/state and 3; focus on, reasons for 6–9; hedonic feelings and 2; imprudent 34; interpersonal relations and 7; judgement theories of 23–4, 54–5; justification of 26–9; misguided 60–1, 74; misplaced 60–1, 74; moral 19, 42n1; moral hazards of 49–53; motivation associated with 3; objects of 65–7; rational accessibility of 85–7; reactive 103–8, 113–15, 139–40; reasons for, right and wrong 84–5; recalcitrant 30–33, 44n18, 86; somatic theories of 54, 56, 85; subject and, impact on 9–10; targeted 22, 42, 45, 48, 59, 64–5, 67–8, 74, 134; un-justified 25–30; vagueness about 4; *see also specific type*; unfair emotions
empowerment and responsibility 13–14
episodes, emotional 118–21
epistemic unfittingness 24–25, 44n6
evaluative attitudes 112–13
evaluative control 84
ex ante justification 70
ex ante morality 70–1
expectations: co-valuational 108–15; difference of personal relationships and 111; holding others to 101–4; normative 102–3; predictive 103; values and 112–13
ex post justification 70
ex post morality 70–1
extensional adequacy 57–60

fair emotions 5, 114–15
fairness 5, 45
fair reaction for unfair emotions: anger 114–115, 122–3, 126–8, 134–5, 139; challenge of 117–8; episodes versus dispositions and, emotional 118–21; indignation 115, 122–3;

overview 118; protest, moral 131–5; resentment 78–80, 89, 97, 114–15, 118, 120–2; retribution 128–31; revenge 128–31
fear 1, 3, 12, 14, 19, 23–4, 28, 30, 34–5, 40, 48–50, 54–8, 61, 63–7, 82–3, 85–6, 94, 101, 121
feelings *see* emotions; *specific type*
fittingness 12–13, 25–30, 57, 60, 83, 85, 140
fitting reactive emotions 113–15
Fricker, Miranda 112

guilt 1, 13, 34, 94, 100–101, 104, 129–32, 142

hedonic feelings 2
Hieronymi, Pamela 86, 96
human regard, basic 111–14
Hume, David 112
hypocrisy 22, 29–30, 33, 42

ill will 117–8, 123–6
imprudent emotions 34
inconsistency criticism of emotions 22, 30–3
incorrectness 23–4
indignation 11, 115, 122–3
indirect voluntary control 80, 83–4
interpersonal directedness feature of unfair emotions 48, 52, 64–7
interpersonal relations and emotions 7
intrinsic moral justification 62–3
involuntary expression of emotions 9–10

Jacobson, Daniel 57
James, William 2, 56
judgement sensitivity 85
judgement theories 23–4, 54–5
justification: of emotions 26–9; *ex ante* 70; *ex post* 70; fittingness versus 25–30; intrinsic moral 62–3; moral 62–3; *pro tanto* moral 48, 62, 114, 117, 128

Kant, Immanuel 112

Levine, Michael 50

managerial control 84
mental states 7; *see also* emotions
misguided emotions 60–1, 74
misplaced emotions 60–1, 74
moral accountability 101
moral criticism of emotions 4–5, 22, 37–43
moral emotions 19, 42n1
moral hazards of emotions 49–53
morality 70–1
moral justification, intrinsic 62–3
morally objectionable feature of unfair emotions 1, 47–57
moral obligations 105–6
moral reactive attitudes 104–8
moral status of a person, misrepresenting 29
motivation, emotions associated with 3

non-comparative unfairness 46–8, 64
non-moral attitudes 104–8
non-moral reactive attitudes 104–8
non-reactive attitudes 104
normative expectations 102–3
Nussbaum, Martha 37–38, 49–50, 72n12, 127, 131–3, 140

objects of emotions 65–7

perceptual theories 23–4, 31–2, 85
personal reactive attitudes 110
predictive expectations 103
pro tanto moral justification 48, 62, 114, 117, 128
protest, moral 131–5, 139–40
prudential criticism of emotions 33–6

quality of will 106–8
questions applied to criticisms of emotions: inconsistency 31–3; moral 40–1; prudential 34–6; unfairness 18–21; unfittingness 26–9

rational accessibility of emotions 85–7
rational control 81, 84–5
rational relations 85
reactive attitudes 88–9, 96–7, 101, 104–8, 139–40
reactive attitudes (RA) theory 12, 75, 77–8, 97, 100–101, 103, 117
reactive emotions 103–8, 113–15, 139–40
reappraisal 81

recalcitrant emotions 30–33, 44n18, 86
recognition and anger 131–5
regard, basic human 111–14
regret 21, 33–4, 41, 43, 125
remorse 88, 130–2
resentment: accountability and 96–7, 100–101, 118; aim of 129–30; anger versus 118–23; blame and 75, 99n52, 101, 126; control and 97; co-valuational expectations and 108–10; description of 103; distinctions among anger and indignation and 118–23; as emotional disposition 103, 118–21, 135; expression of 11, 102; fair 78–80, 89, 97, 114–15, 118, 120–2; as ill will 117–18, 123–6; indignation versus 118–23; misplaced 37, 138; motivational profile of 126; punishment and 77; quality of will and 106–7; reactive attitudes and 96–7, 101; as reactive emotion 103; responsibility and 78, *78*; for unfair emotions 138
responsibility: for anger 92; blame and 90; control and 75–81, *78*, 97n4; downsides of taking 13; empowerment and 13–14; rational 75, 81, 84–5; reactive theory of moral 12, 15, 142; resentment and 78, *78*; voluntary 75, 80–4; *see also* accountability
retribution 37, 50, 96, 118, 128–33
revenge 37, 50–1, 117–18, 127–33, 140

sanctions, social 88–9, 93–6
Scanlon, Thomas M. 77, 112
schadenfreude 61
second-personal moral criticism of emotions *see* unfairness criticism of emotions
second-personal sentiment 103; *see also* emotional dispositions
sentiment *see* emotions; *specific type*
Silva, Laura 131, 140
Smith, Angela M. 76, 85–6, 92, 96–7, 131
social sanctions 88–9, 93–6
somatic theories 54, 56, 85
Strawson, Peter F. 102, 106–8, 111, 113, 123
superbia 38–40

targeted action tendencies 67
targeted emotions 22, 42, 45, 48, 59, 64–5, 67–8, 74, 134
Teroni, Fabrice 65, 118
thought policing 142

uncertainty 138–9
unfair emotions: accountability for 140; benefits of theoretical account for 12–14, 141–2; challenges with 10–12; conditionality feature of 48, 57–63; debunking strategy and 6, 9; distinction from other criticisms and 137; examples of 45–6; interpersonal directedness feature of 48, 64–7; morally objectional feature of 1, 47–57; negative involvement of other people and 34; non-comparative unfairness and 46; overview 14–15; phenomenon of 4–10; philosophical approaches to 5–6; problems of 10–2; resentment for 138; responses to 1; thesis, main 1; uncertainty and 138–9; wrongs and 11; *see also specific type*
unfairness criticism of emotions: comparative fairness and 45, 47; comparative unfairness and 45–6; definition of 4; directedness and 42, 66; examples of 4–5, 17–8; moral criticism versus 4–5, 39, 42–3; non-comparative unfairness and 46–8; other criticisms versus 18, 42–3; outlook of 138–41; philosophy and, moral 137; questions applied to 18–23; sensitivity and 143; term for, another 72n1; unfittingness versus 30; wrongs/wronging and 11, 68
unfittingness criticism of emotions: description of 23–4; epistemic 24–25, 44n6; examples 23; inconsistency criticism versus 33; incorrectness and 23–4; questions applied to 26–9; unfairness criticism of emotions versus 30; un-justified emotions and 25–30
un-justified emotions 25–30

values and expectations 112–13
voluntary control 75, 80–4, 94

Wallace, R. Jay 102, 105–6, 113
Watson, Gary 89
will 106–8, 117–18, 123–6
wrongs/wronging: accountability and 70–2; claims and 68–70; directed 67–72, 74–5; directedness of moral 11; direction of 70–1; unfairness criticism of emotions and 11, 68