Introduction

Andreas Brekke Carlsson

Self-blame is an integral part of our lives. We often blame *ourselves* for our failings, and we experience a familiar set of unpleasant emotions such as guilt, shame, regret, or remorse. Self-blame is also often what we aim for when we blame *others* – we want the people we blame to recognize their wrongdoings and blame themselves for it. Moreover, self-blame is typically considered to be a necessary condition for forgiveness. If the wrongdoer has not blamed herself for her action, say by experiencing guilt or remorse, forgiveness may seem inappropriate. Yet so far, self-blame has not been an integral part of the theoretical debate about the nature of blame and its relation to moral responsibility. This volume seeks to remedy this omission.

Until recently, philosophers working on blame and moral responsibility have focused almost exclusively on other-directed blame. In the Strawsonian tradition, the emphasis has been on anger, resentment and indignation, and the communication of these emotions (Wallace, 1994; Watson, 1996; McKenna, 2012; Macnamara, 2015a; Shoemaker, 2017). Alternative views have seen blame as a way of modifying one's relation with others (Scanlon, 2008), as a belief-desire pair (Sher, 2006), or as an expression of protest (Hieronymi, 2001; Smith, 2012; Talbert, 2012). It is unclear how self-blame fits into these accounts of the nature of blame. On the face of it, self-blame is not obviously a form of communication, a modification of relationships, or the expression of protest.

It is also unclear how a focus on self-blame will affect accounts of moral responsibility. An agent is blameworthy to the extent that it would be appropriate to blame her. The question of what makes it appropriate to blame other people is arguably the central issue in the literature on moral responsibility. But the question of what would make self-blame appropriate has received far less attention. Self-blame thus raises new questions. It also makes old questions take on a new significance. One of those is the question of whether wrongdoers can deserve pain or harm for their

wrongdoings. On many accounts, self-blame is intrinsically unpleasant or painful, for example if self-blame is identified with guilt. This raises the question of whether and how this particular form of pain can be justified. For example, is the pain a mere side effect of a fitting recognition of one's own wrongdoing, or is this pain something that wrongdoers deserve in the sense that it is noninstrumentally good and just that they experience it?

The contributions to this volume show that thinking carefully about self-blame might change or challenge our perspectives on traditional problems in the debate on blame and moral responsibility and open new avenues for research in moral philosophy, moral psychology, and the philosophy of punishment.

The issues that will be discussed in the chapters that follow can be divided into three groups of questions that intersect in interesting ways.

(1) The *nature* of self-blame: There are many competing accounts of other-directed blame. How does self-blame fit into a comprehensive account of blame? It seems that we blame ourselves both for our moral and nonmoral failures. What is the difference between moral and nonmoral kinds of self-blame? It is common to understand moral self-blame as the emotion of guilt. If so, what is the nature of guilt? Is it possible to blame oneself without experiencing guilt? The nature of self-blame is importantly connected to its justification. Our reasons for blame are often backward-looking: We tend to think it is appropriate to blame an agent simply in virtue of what they have done. However, sometimes our justifications are forward-looking. We blame in order to influence, educate, or generate other beneficial consequences. Can there be forward-looking reasons for self-blame and is it possible to develop an account on which self-blame is only justified by forward-looking considerations?
(2) The *ethics* of self-blame: There is a thriving debate concerning the norms of other-directed blame. But what are the norms governing self-blame? Is self-blame something that we should express to those who we have wronged, or should it rather be suffered in silence? There are interesting asymmetries between the normative expectations concerning blaming oneself and others. For example, there are many cases where it may seem appropriate to blame ourselves, but where it is less clear whether it is appropriate for others to blame us. Are there different standards for blaming oneself than for blaming others? Finally, we often experience emotional reactions that are at

odds with our own evaluations and judgments. Are these reactions something for which we should blame ourselves?

(3) The relation between self-blame and theories of *moral responsibility*: Given that self-blame and other-directed blame differ in many respects, which of them should be fundamental in our conception of blameworthiness? Many theories on moral responsibility focus on the communicative aspects of blame. How does self-blame fit into this picture? Other theories emphasize that blame is sometimes harmful. As a result, it is often assumed that agents must fulfil certain conditions for other-directed blame to be justified. Does the fact that guilt is intrinsically painful support a strict control condition on moral blameworthiness and is the painfulness of guilt something we deserve?

These three topics correspond with the three parts of the volume.

Part I concerns the nature of self-blame. What is it to blame oneself? According to a traditional view, self-blame is identified with the emotion of guilt. In their chapter, Justin D'Arms and Daniel Jacobson develop an account of what guilt is and how it relates to moral responsibility. The Strawsonian approach to responsibility tries to explain what it is to be morally responsible for one's actions in terms of being an appropriate object of the reactive attitudes, such as resentment, indignation, and guilt. How informative such explanations can be depends in part on whether an adequate characterization of these attitudes can be given without appeal to the concept of responsibility itself. Guilt, D'Arms and Jacobson argue, is the most promising candidate for a Strawsonian account of the first-personal case of holding oneself morally responsible. The question is whether guilt can be characterized without appeal to responsibility or any similar concept. D'Arms and Jacobson argue that it can. This chapter offers a theory of guilt as a motivational state involving a goal and specific action tendencies that constitute direct and urgent means of meeting that goal. Despite its cognitive complexity, guilt is like simpler emotions such as anger and fear in how its goals and action tendencies are discontinuous with practical reasoning. The motivational theory of guilt, developed in this chapter, provides an important tool for theorizing about first-personal responsibility practices in Strawsonian terms.

If we focus on moral wrongdoing, it seems natural to think of self-blame as guilt. But once we consider self-blame for nonmoral failures, the picture may look very different. The starting point of David Shoemaker's contribution is *athletic* self-blame. When Tom Brady throws an interception, he

yells at himself and pounds his fists on his helmet. When Serena Williams misses a shot, she breaks her racket. These athletes, Shoemaker argues, are clearly blaming themselves. Surprisingly, though, current theories of blame have a hard time accounting for such cases. Most theories of blame take other-blame as their paradigm, typically thought to be a response to poor quality of will or moral wrongdoing and consisting in some kind of relationship modification, communication, or protest. None of these features seem to apply in the cases of athletic self-blame. Recently, some theorists have taken self-blame to be a more fundamental paradigm than other-blame. But they are focused on self-blame as guilt, which again, according to Shoemaker, can't capture the athletic cases, because Tom Brady and Serena Williams are not feeling anything like guilt when they blame themselves. In reply to these problems, Shoemaker offers a new theory of self-blame, one that takes athletic cases as its starting point. He draws on recent interesting psychological work on the phenomenon of *self-talk* to make the case that the emotional core of self-blame is in fact very different than that of other-directed blame.

Douglas Portmore, by contrast, develops a unified account of self-blame and other-directed blame. Portmore's goal is to provide a *comprehensive* account of blame, on which deserved guilt, regret, and remorse play an integral part. It is widely noted that blame is multifarious. It can be passionate or dispassionate. It can be expressed or kept private. We blame both the living and the dead. And we blame ourselves as well as others. What is more, we blame ourselves, not only for our moral failings but also for our nonmoral failings: for our aesthetic bad taste, gustatory self-indulgence, or poor athletic performance. And we blame ourselves both for things over which we exerted agential control (e.g., our voluntary acts) and for things over which we lacked such control (e.g., our desires, beliefs, and intentions). Portmore argues that, despite this manifest diversity in our blaming practices, it is possible to provide a comprehensive account of blame. Blame, according to Portmore, essentially involves representing the wrongdoer as not having experienced all the guilt, regret, or remorse the wrongdoer deserves. Based on this idea, he proposes a set of necessary and sufficient conditions of blame and argues that this proposal has a number of advantages over competing theories.

Portmore's account of blame and self-blame presupposes a notion of deserved guilt, regret, and remorse. In his chapter, Derk Pereboom develops a notion of self-blame that *does not* invoke the notion of deserved pain or harm. In previous work, Pereboom has argued that causal determinism and the absence of control that indeterminism implies will

undermine this kind of desert. In this chapter, he focuses on developing a *nonretributive* conception of other-directed blame and self-blame. On this account, to blame is to take on a nonretributive stance of moral protest. The reasons for moral protest are forward-looking: moral formation or reconciliation in a relationship that has been impaired due to wrongdoing, protection from wrongdoing, and restoration of the integrity of those who were wronged. To blame oneself, according to Pereboom, is to take on a stance of moral protest toward oneself in virtue of an action one regards as morally wrong. The reasons one has for doing so are forward-looking and include moral formation and reconciliation in a relationship that has been impaired as a result of one's wrongdoing. Which emotional reactions would aptly accompany this form of self-blame? According to Pereboom, guilt presupposes desert. But *regret*, that is, a painful response to one's own wrongdoing, does not involve the supposition that the pain it involves is basically deserved and can thus be apt even if no one deserves to experience pain or harm.

Part II concerns the ethics of self-blame. Are there different standards for blaming oneself than for blaming others? Dana Nelkin begins her contribution by observing that there is a striking asymmetry in our normative expectations of degrees of self-blame and degrees of other-directed blame. There are many situations in which it seems intuitively plausible that a person should blame herself to a certain degree, while at the same time, it is also appropriate for others to blame her to a lesser degree. This calls out for explanation. Nelkin canvasses the prospects for rejecting the idea that there is any systematic explanation to be found. She also critically discusses a variety of possible explanations that purport to justify a genuine asymmetry between the norms of self-blame and other-directed blame. The latter group includes explanations according to which it is a virtue to over-blame in one's own case and in which it is a virtue to be disposed to under-blame in the case of others. Instead, Nelkin argues that a central and systematic explanation relies in part on a general moral principle according to which asymmetric risk imposition between self and others is justified. Nelkin concludes by exploring the implications of this view for whether we should privilege intuitions about self-blame, other-directed blame, or neither in philosophical theorizing.

Self-blame may but need not be expressed. *Should* self-blame be expressed, and if so under what conditions? Hannah Tierney's chapter develops an important norm for the expression of self-blame that she calls "Don't Suffer in Silence." When we blame ourselves, we ought not do so privately. Rather, we should, *ceteris paribus*, express our self-blame to those

we have wronged. Tierney centers her discussion around a paradigmatic form of expressing self-blame: expressing guilt. She notes several important reasons for expressing one's guilt. Such expression does important interpersonal work. In confessing our guilt to those we have wronged, we begin the process of repairing our relationships with them. Expressions of guilt can ease victims' suffering, restore something important that they have lost (or that was taken from them) and reaffirm their standing in the moral community. But such expressions can also serve as an ameliorative function for the wrongdoer. A failure to express one's guilt can make the wrongdoer suffer more guilt than she deserves. Tierney ends her chapter by exploring how the "Don't Suffer in Silence" norm can contribute to our understanding of the ethics of self-blame as well as the nature of blameworthiness itself.

Interesting normative questions concerning self-blame arise in cases when we evaluate our own emotional responses. Krista Thomason's chapter explores an important moral experience: that of judging ourselves for our emotional responses. Often the emotions that we criticize are *recalcitrant*: they are emotions that we do not endorse or that conflict with our considered judgments. Thomason notes that most of the philosophical literature on recalcitrant emotions focuses on whether and how they are possible or whether and how they are irrational. Thomason focuses instead on the ways we blame ourselves for recalcitrant emotions. She argues that it is harder than it looks to explain self-blame for recalcitrant emotions. Recalcitrance alone does not give us a reason to feel any particular way about our emotions, and it does not provide sufficient grounds for self-blame.

Part III investigates the relationship between self-blame and moral responsibility. In his chapter, Michael McKenna examines the role of both self-blame and guilt within the context of his conversational theory of moral responsibility. According to McKenna's own theory as well as communicative theories of responsibility more generally, the central examples of blame involve others overtly and directly blaming the one who is blameworthy and so communicating with the culpable party. Some philosophers have recently placed guilt and self-blame at the heart of moral responsibility's nature. They also have made the deservingness of guilt the most fundamental normative consideration in justifying the harms of blaming. Doing so appears to threaten conversational and other communicative theories of moral responsibility. In response, McKenna argues that guilt and self-blame cannot play the fundamental grounding role in a theory of moral responsibility. As a result, conversational and

other communicative theories are not in jeopardy. Rather, what is required is a proper appreciation of the aim and norms of our blaming practices wherein guilt and also self-blame are meant to fit as responses to the blame of others as well as oneself. Along the way, McKenna also argues that self-blame and guilt are distinct things. While it is natural to think that to experience guilt just is to blame oneself, this, according to McKenna, is not so. Although the two are tightly connected, the relationship is nevertheless contingent; one can blame oneself without experiencing guilt, and one can experience guilt without blaming oneself.

In my own chapter, I focus on an often overlooked aspect of blameworthiness. The literature on moral responsibility is ripe with accounts of what it takes for an agent to become blameworthy. By contrast, very little has been written about what it takes for an agent's blameworthiness to cease or diminish. It seems that there are certain things a wrongdoer can feel or do that might make her less blameworthy than she would otherwise have been. She might experience guilt, atone, apologize, and make reparations. I argue that prominent accounts of blameworthiness are unable to explain how such actions and emotions can influence one's blameworthiness. I then present an alternative account. If we understand blameworthiness in terms of deserved guilt rather than fitting resentment, we can give a plausible account of how blameworthiness can change over time. The fact that a wrongdoer has already experienced guilt, atoned, or apologized will make her less deserving of guilt and therefore less blameworthy.

Gunnar Björnsson's chapter also concerns the connection between guilt, desert, and blameworthiness. Central cases of moral blame suggest that blame presupposes that its target deserves to feel guilty and that if one is blameworthy to some degree, one deserves to feel guilt to a corresponding degree. This, some think, is what explains why being blameworthy for something presupposes having had a strong kind of control over it: only given such control is the suffering involved in feeling guilt deserved. Björnsson argues that all this is wrong. By considering a wider range of cases, Björnsson proposes that blame does not presuppose that the target deserves to feel guilt and does not necessarily aim at the target's suffering in recognition of what they have done. In addition to that, he offers an explanation of why, in many cases of moral blameworthiness, the agent nevertheless deserves to feel guilt. The explanation builds on a general account of moral and nonmoral blame and blameworthiness and a version of the popular idea that moral blame targets agents' objectionable quality of will.

We often feel guilty for our wrongdoings. But do we have any reason to feel guilty, and can the pain of guilt be deserved? In the final chapter of this volume, Randolph Clarke and Piers Rawling argue that the facts that make an agent blameworthy also provide the agent with a reason to feel guilty. For example, the facts in question might be that the agent acted freely, with knowledge that the action was wrong, and was moved by ill will. The same set of facts that makes an agent blameworthy also suffices for the agent to deserve to experience the painful emotion of guilt. Clarke and Rawling argue that desert is essential to moral responsibility, that it can be permissible to induce a feeling of guilt in people who are blameworthy, and that it is noninstrumentally good that people who are blameworthy are subject to a fitting feeling of guilt.

PART I

The Nature of Self-Blame

CHAPTER 1

The Motivational Theory of Guilt (and Its Implications for Responsibility)

Justin D'Arms and Daniel Jacobson

The Strawsonian approach to responsibility tries to explain what it is to be morally responsible for one's actions in terms of being an appropriate object of the reactive attitudes (see Strawson, 1962).[1] In order to succeed, the approach must first explain what the relevant attitudes are and what is meant by appropriateness. Although there are both negative and positive reactive attitudes, corresponding to blame and praise, most of the discussion following Strawson focuses on the negative side. It can therefore only hope to capture blameworthiness, not responsibility in general, since to be morally responsible in a good (or neutral) way is surely not to be the appropriate object of a negative attitude. We, too, will focus on blameworthiness, which Strawsonians hope will provide the foundation for a general theory. This chapter develops and answers an important challenge to any such account of responsibility, whatever the reactive attitudes to which it appeals. Our discussion centers on guilt, for reasons to be explained, and hence specifically concerns self-blame. A similar problem arises for other-directed blame, which will require an analogous solution.

The challenge facing the Strawsonian project also faces the sentimentalist project we have been developing for some time, and we will suggest that the same solution applies to both cases. *Sentimentalism*, as we understand it, refers to those views that explain (at least some) values in terms of the emotions; and our own view, *rational sentimentalism*, does so specifically in terms of the fittingness of emotions – or, equivalently, of what merits them – where merit and fit are understood to be notions of correctness. We have argued that considerable confusion arises from the failure to differentiate between fittingness and other forms of appropriateness.[2]

[1] There are other ways to read Strawson's classic paper, but this is what we shall mean in referring to the Strawsonian tradition. McKenna (2012), Rosen (2015), Shoemaker (2017), and Wallace (1994), among others, are all Strawsonians in this sense.

[2] See D'Arms and Jacobson (2000) for more on differentiating such notions of appropriateness.