

Blameworthiness as Deserved Guilt

Andreas Brekke Carlsson

The final version is available at <https://link.springer.com/article/10.1007/s10892-016-9241-x>

1. Introduction

It seems unfair to blame someone for something over which he had no control. In line with this popular view, many writers accept a control condition on moral blameworthiness:

CONTROL: Agents are only blameworthy for that over which they have either direct or indirect¹ control.

Recently, however, several writers have argued that we can be blameworthy for occurrences that appear to be outside our control, such as attitudes, beliefs and thoughtless omissions.² This has put pressure on the condition and prompted the question of *why* CONTROL should be a requirement on moral blameworthiness. The most natural way of answering this question relies on the notion of blame.³ Being blameworthy is to be worthy of blame. Following P. F. Strawson (1962), many writers have understood blame in terms certain negative emotional responses — the so-called reactive attitudes. The negative reactive attitudes include resentment, indignation and guilt, though in fact almost every writer in this tradition privileges resentment and indignation. Since this kind of blame is unpleasant or painful to its recipients, it would be unfair or inappropriate to blame an agent for something she cannot

¹ Sometimes one is blameworthy even though one lacks *direct* control over one's action. If I drive while being drunk I am not in direct control of my actions. But suppose I was in control when I started drinking and knew that I would drive the car home. Hence I might be blameworthy for my drunk driving in virtue of having *indirect* control over that action

² See for example Smith (2005), Scanlon (2008) and Sher (2009).

³ For a critical discussion of other ways of justifying CONTROL, see Sher (2009: ch.4).

control.⁴ This popular idea⁵ has a strong intuitive appeal. In general, it seems to be unfair to impose adverse or unwelcome treatment for something that is outside one's control. Nevertheless, this line of reasoning has come under attack. Pamela Hieronymi (2004) and Peter Graham (2014) have argued that the above argument misconstrues the nature of blame.⁶ The argument presupposes that blame is harmful. But once we consider what blame consists in, it is not obvious that this is the case. After all, resentment and indignation need not be expressed and when these emotions are expressed the wrongdoer might not care. But if blame is not harmful, why should we care about control?

This paper aims at developing a plausible conception of blameworthiness, according to which blameworthiness is necessarily connected to suffering and to use this account to explain *why* CONTROL is a necessary condition of blameworthiness.

I will argue that blameworthiness should not simply be understood in terms of the appropriateness of the blame of others. Instead, it should be understood in terms of the appropriateness of the wrongdoer *blaming himself*. Being blameworthy, I will argue, simply is to deserve guilt. This account is inspired by two recent papers by Randolph Clarke (2013; 2016). Clarke defends what he calls the *minimal desert thesis*: Agents who are blameworthy deserve to feel guilty because they are blameworthy. His thesis therefore presupposes a prior understanding of what it is to be blameworthy. I will argue that we should reverse the order of explanation. Agents are blameworthy *because* they deserve to feel guilty. The advantage of this reversal is twofold: It will allow us to develop a new account of what it *is* to be blameworthy and to explain why the conditions of blameworthiness are as they are.

⁴ There are other notions of blame that are even less supportive of CONTROL. Arpaly (2003) takes blame to be a judgment about quality of will, Sher (2006) understands blame as a belief- desire pair and Scanlon (2008) gives an account of blame as a modification of the intentions and expectations that constitute a relationship. All of these accounts capture important aspects of what we call blame. But I am interested in the notion of blame that is at issue in the debate about free will and that can be challenged by sceptical arguments. As Pereboom (2013: 191) has pointed out concerning Scanlon's account: a free will sceptic can *accept* that an action often shows something about the agent's attitudes that impairs the relations others can have with him or her. Similar points can be made about the notions of blame developed by Arpaly and Sher.

⁵ See for example Wallace (1994), Watson (1996/2004), Rosen (2002) and Nelkin (2011: ch. 2).

⁶ Scanlon (2008: 198) raises a similar criticism.

Contrary to what the critics claim, there is a necessary connection between blame and suffering. Being blamed by someone else is not necessarily harmful to the wrongdoer. But to blame oneself - to feel guilty- necessarily involves suffering. The pain of guilt is constitutive of feeling guilt: if it doesn't hurt, it isn't guilt. With this conception of blameworthiness we can explain why the control condition should obtain. We are morally blameworthy for A only if A was (directly or indirectly) under our control because (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.

The main task of the paper is to argue for (a). Condition b), I hope, needs no defence. I take (c) to have a strong intuitive appeal, but I will not provide any arguments for it. Control can be understood in many different ways. I will not try to specify the relevant control condition in this paper. Instead I will provide a framework for how we should *address* this question: the relevant control condition is the condition required for agents to deserve guilt.

I will proceed as follows. In section 2, I will briefly present the reactive account of blame, according to which blameworthiness is understood in terms of reactive attitudes. On the reactive account, an agent is blameworthy only if he is an appropriate target of the reactive attitudes. This idea can be spelled out in different ways that have different implications for the plausibility of CONTROL. In section 3, I discuss the claim that agents are blameworthy to the extent that it would be appropriate to *feel* resentment and indignation towards them. On this view, there is no harm involved in blaming: Blame is an emotion, not a sanction. I argue that this conception is unsatisfactory because it fails to make sense of our practice of expressing blame. In section 4, I discuss the claim that agents are blameworthy to the extent that it would be fair to *express* the reactive attitudes towards them, or that they would deserve to be the recipient of expression of the reactive attitudes. On these views, there *is* a harm involved in blaming, but the connection between expressing blame and harm is

indirect and contingent. This, I argue, leads to some counterintuitive results. In section 5, I present my own view: an agent is blameworthy in virtue of the fact and to the extent that he deserves to feel guilt. Here the relevant notion of blame is understood as an emotion, not as any form of overt treatment. But this conception of blameworthiness nevertheless necessarily involves suffering. This is, however, a special kind of suffering: the pain of recognizing what you have done. I will compare my account with competing views and discuss the difference between my conception of blameworthiness and Clarke's minimal desert thesis. In section 6 I will show how my conception can solve the problems discussed in sections 3 and 4. I will then use this conception of blameworthiness to give a tentative argument for CONTROL. Section 7 responds to an objection; section 8 concludes.

2. The Conception of Blameworthiness

What is it to be blameworthy? Following R. Jay Wallace (1994),⁷ I will use the following generic schema as a starting point:

Agent S is morally blameworthy (for action X) if and only if it is appropriate to blame S (for X).

I will also follow Wallace (1994) and Rosen (2015: 73) in taking the explanatory priority to go from right to left in this biconditional. When S is blameworthy for X, S is so in virtue of the fact that it is appropriate to blame S for X. The conditions of blameworthiness are as they are because for S to be blameworthy for X just is for blame of S for X to be appropriate.

In order for the above schema to be informative, we need to specify the terms: what is “blame,” and what is it for blame to be “appropriate”? In this paper I will explore several

⁷ This way of understanding blameworthiness has been adopted by a number of philosophers. See for instance Fischer and Ravizza (1998), McKenna (2012), Brink and Nelkin (2013), Pereboom (2013) and Rosen (2015).

different versions of the reactive account, according to which blame is understood as negative reactive attitudes: indignation, resentment and guilt. We feel resentment when the wrongdoing is directed towards ourselves, and indignation when the wrongdoing is directed against others. When we feel guilt, we perceive ourselves as wrongdoers. These reactive attitudes have a rebuking character, which set them apart from contempt and shame (Graham: 2014: 389). Resentment and indignation are angry emotions, and guilt is intrinsically painful. (Wolf 2011; Rosen 2015).

Using the reactive account of blame, we can narrow down our conception of blameworthiness:

Agent S is blameworthy (for X) if and only if it would be appropriate to respond to S's X-ing with the reactive attitudes: resentment, indignation or guilt (in the reflexive case).

This formulation leaves it open whether we should understand "to respond" in terms of *feeling* or *expressing* the reactive attitudes. It also lacks a specification of the relevant sense of "appropriate." How we specify these terms will have different implications for the plausibility of CONTROL. In the next two sections, I will discuss how two different versions of the reactive accounts deal with these issues.

3. Blame as Emotions

If blame is identified with the reactive attitudes, it seems natural to understand blameworthiness in terms of the appropriateness of *experiencing* these emotions (as opposed to expressing them). Pamela Hieronymi (2004) and Peter Graham (2014) have defended this view. Graham gives the following definition, which I will call APPROPRIATE EMOTIONS (AE):

X is blameworthy for ϕ -ing iff_{def} the feeling of a blame emotion—resentment, indignation, or guilt—toward X in response to ϕ -ing would be appropriate.

(Graham 2014: 390)

How should we understand the notion of appropriateness involved here? One possibility is that the appropriateness of blame should be understood in terms of moral norms, such as fairness (Wallace 1994; Watson 1996/2004). This would make sense if blame were a sanction. If blaming someone entailed some form of adverse treatment, this activity would plausibly be governed by norms of fairness or desert. But both Graham and Hieronymi emphasize that blaming is not a form of sanction. “As blaming someone consists in feeling certain emotions toward her, it is not a form of adverse treatment at all” (Graham 2014: 391). On AE, the relevant notion of blame is understood in terms of private, emotional reactions, and emotions themselves are not plausibly governed by moral norms, according to Graham and Hieronymi.

Both Graham and Hieronymi argue instead that the relevant norm is *correctness*. According to Hieronymi, reactive attitudes are responses to certain judgments. These attitudes are appropriate when the judgments are correct. Similarly, Graham argues that emotions like resentment and indignation are constituted by thoughts. He draws an analogy between blame and fear. When John fears a tiger, he imputes to the tiger the property of being dangerous to him. Fear of tigers has a propositional content: that tigers are dangerous. This emotion is appropriate just in case its propositional content is true. Fear of tigers is appropriate (in certain circumstances); fear of kittens is not. The same structure, Graham argues, applies to resentment and indignation. These emotions impute certain features to its objects. They are appropriate just in case the object actually has these features.

Which thoughts or judgments are involved in the reactive attitudes? Both Graham and Hieronymi argue that the judgments concern the agent’s *quality of will*: Hieronymi speaks of

“ill will,” while Graham prefers “violations of a moral requirement of respect.” Indignation and resentment are responses to the perceived ill will of an agent. These emotions are appropriate only if the agent actually displayed ill will. This makes intuitive sense: we typically feel the reactive attitudes when we encounter malign intentions or a lack of proper concern for others. The point of excuses is often to show that one’s action, contrary to appearances, did not express any ill will. If John says that he did not mean to step on Mary’s toes, he wants to forestall the assumption that he wanted to hurt her. If Mary says she did not know that John was ill, she wants to assure that it was not because of any lack of concern for his well-being that she did not call.

Most philosophers would agree that ill will is an important, and perhaps necessary, condition on being blameworthy. But Graham’s account seems to suggest the much stronger claim that ill will is a *sufficient* condition on blameworthiness. He concludes his paper with this summary:

(1) To be blameworthy for ϕ -ing is to be such that it would be appropriate for a blame emotion to be felt toward one on account of one’s ϕ -ing. (2) The blame emotions are appropriately felt just in case the propositional content of those emotions is true. And (3) the content of a blame emotion felt toward a person for ϕ -ing is that, in ϕ -ing, that person has violated a moral requirement of respect.

(Graham 2014: 408)

This claim seems to be too strong. A small child or a demented adult appears able to violate moral requirements of respect, but this does not mean that it would be appropriate to feel the reactive attitudes when they do so. Such agents appear to be exempted from appropriate blame. Graham might perhaps solve this problem by supplementing the propositional content

of the blame emotions. He could for example add a requirement of general capacities, or moral competence. I will return to this issue in section 5.

Having presented AE, I will now consider how this conception of blameworthiness relates to CONTROL. Even if Graham's account were amended with a requirement of general capacities or moral competence, AE might still be incompatible with CONTROL. Morally competent adults can display ill will in actions, attitudes and omissions over which they do not have control. Consider one of Graham's examples:

Torture: Adam passes by a window into a room in which Sam is being tortured.

Sam sees Adam looking in and is aware that Adam knows that he cannot intervene and help him. Sam also notices, however, that as his own agony increases a smile of sheer delight creeps across Adam's face. At this sight, resentment toward Adam wells up within Sam. (Graham 2014: 396)

Graham emphasizes that Adam's smile is not something over which he has control: "Like blushing, it is more of a bodily occurrence than an action" (ibid). Yet Graham claims that Sam's resentment of Adam is entirely appropriate, that "there is no sense of inappropriate according to which it is inappropriate" (ibid). I think many would share Graham's assessment that it is appropriate for Sam to feel resentment towards Adam. Still, since Adam could not control his smile, one might also take it to be unfair to sanction him by expressing resentment or indignation, or to harm him in any way. But these intuitions concerning the fairness of sanctions would be misplaced, given Graham's account of blame; on AE, blame does not entail any treatment at all.

The potential conflict between AE and CONTROL can be illustrated with other examples from the literature:

Moral ignorance: A Hittite lord beats his slaves. He believes that he is entitled to do so as they are his property. He thus acts on a false moral belief. But it is plausible that he is not blameworthy for his ignorance. The slaveholder lived in a period when the legitimacy of chattel slavery was taken for granted and it would be extremely difficult for a man of that time to figure out that the practice is wrong. (Adapted from Rosen 2002: 64)

Manipulation (Plum kills White): Plum is like an ordinary human being, except that a team of neuroscientists has programmed him at the beginning of his life to weigh reasons for action so that he is often but not exclusively rationally egoistic, with the consequence that in the circumstance in which he now finds himself, he is causally determined to undertake the reasons-responsive process of deliberation and to possess the set of first- and second-order desires that result in his killing White. (Pereboom in Fischer et al. 2007: 95-96)

In these scenarios, the wrongdoers act with ill will, or, to use Graham's terminology, violate a requirement of respect. A proponent of AE might therefore say that it would be perfectly appropriate for the slave to feel resentment towards the slaveholder or for White's relative to feel indignation or resentment towards Plum. But it is also plausible that these agents manifest ill will because of factors over which they had no control. Neither the Hittite Lord's belief in the legitimacy of chattel slavery, nor Plum's egoistical reasoning is under their control.⁸ Because of this lack of control, it has been argued that it is unfair to blame them. Graham and Hieronymi may reply to this objection as follows: when people have the intuition that the wrongdoers in these scenarios are not blameworthy, they have *a different conception of blameworthiness* than AE in mind. This is supported by the literature. When Rosen argues

⁸ Zimmerman (1997) and Rosen (2002) argue that moral ignorance cannot be the locus of original responsibility because we are not in direct control of our belief formation.

that the Hittite lord is not blameworthy, he explicitly states that blame is a sanction: “Even when it is not expressed, it is a form of adverse treatment: a form of psychic punishment. And like any adverse treatment, blame is governed by moral norms” (Rosen 2002: 73). Pereboom argues that Plum is not blameworthy for killing White. But according to Pereboom, an agent is blameworthy to the extent that “she would deserve to be *the recipient of an expression of moral resentment or indignation*” (Pereboom 2013: 189; my emphasis). Both of these arguments rely on considerations based on fair or deserved *sanctions*. On AE, however, blame is not a sanction and being blameworthy does not entail any form of harm or adverse treatment. If AE is correct, defenders of CONTROL will be unable to employ the most effective and straightforward argument for the principle: that it would be unfair or undeserved to suffer any harm for something you cannot control. Does this mean that CONTROL is based on a misconception of what blameworthiness is?

I do not believe so. I will now argue that we have independent reasons to reject AE. Our interest in the issue of moral responsibility is to a large degree motivated by our interest in whether our blaming practices are justified. An account of blameworthiness should reflect this. If we know that an agent is blameworthy, we should also know that it is *pro tanto permissible* to express blame. Let us call this the *practicality requirement*. The requirement is *pro tanto* because there will often be features of a situation that can defeat the permissibility of expressing blame. Sometimes we lack the moral standing to express blame, and sometimes it would lead to disastrous consequences. If an agent is suicidal, it may be wrong to express our resentment.⁹ However, if an account of blameworthiness does not give us any way of understanding whether our blaming practises are justified, it fails to answer the question that gives the debate about blameworthiness its moral urgency.

⁹ See Smith (2007) for an interesting discussion of such cases.

AE tells us that an agent is blameworthy to the extent that the content of the reactive attitudes is correct. This will provide us with *some* conditions on when it would be appropriate to express blame: it will usually be a mistake to express an attitude you do not have, because it is disingenuous. And it will usually be a different kind of mistake to express an attitude you have, but that is not in fact correct, because you will be expressing an attitude whose content is false. But this is not sufficient to show that it is pro tanto appropriate to express the reactive attitudes. Joel Feinberg (1970: 82) suggests that outward expressions of blame could be justified because they are “natural or conventional means of expressing the morally fitting attitudes.” But once the reactive attitudes are expressed, they will take the form of reproach, scolding, remonstrance, etc., and these actions can be harmful to their recipients. If the activities are harmful, they should be regulated by norms of fairness, or be something that the wrongdoer deserves. But desert or norms of fairness are not built into AE. As we have seen, appropriateness is not understood as a moral notion, but rather one of correctness: blame is appropriate if the judgment of ill will is correct. This does not licence any harm towards the wrongdoer. Indeed, the very reason for adopting correctness and not fairness or desert as the relevant interpretation of appropriateness was that blame in itself, according to Graham (2014: 391-392), is not harmful. The fact that an agent is blameworthy, on this account, will therefore provide us with no justification to express the appropriate blame.¹⁰ AE fails the practicality requirement.

¹⁰ Of course, there can be other justifications for expressing the reactive attitudes. Blame also has a *forward-looking* aspect, and scolding and reproaching may sometimes have beneficial consequences. We sometimes express blame in order to make people understand that they have done something wrong. This might inspire a change of heart, new moral insights, and regret. Even when the prospects of moral reform appear dim, blame might be justified as a form of moral protest. As Matthew Talbert (2012: 105-107) has argued, blame can be a way of affirming one’s moral standing. When confronted by the ill will of a committed racist or a psychopath, your expression of blame may not influence the wrongdoer, but it may still be valuable as a “defiant declaration” of moral standing and moral standards for “the protestor and his fellow sufferers” (107). There are thus several plausible justifications for expressing blame. But these justifications are independent of the wrongdoer’s *status as blameworthy*. We may have reasons to influence or to affirm our moral standing towards people who are not blameworthy at all. Of course, one could construct consequentialist versions of *blameworthiness* according to which agents are blameworthy to the extent that blaming them would produce good consequences or affirm the victim’s moral standing. But such accounts would be highly revisionary and a far cry from AE.

Perhaps this is too quick. Adherents of AE might argue as follows: the blame emotions are reactions to judgments about a wrongdoer's quality of will. When we are expressing blame, we are in effect making this judgment explicit. But when the judgment is correct—when the wrongdoer manifested ill will—there is nothing *unfair* about expressing this judgment. Expressed blame is simply a way of conveying to the wrongdoer that he acted with ill will. Hieronymi suggests an argument along these lines:

Compare resentment with distrust. Distrust also seems to carry a special, characteristic burdensome force. [...] Yet it is not clear that those who interact with [the untrustworthy person] can be charged with unfairness in distrusting her. [...] The distrust simply marks the fact that a person's unreliability or untrustworthiness is being known. Something similar, I will suggest, is true of the reactive attitudes resentment and indignation. (Hieronymi 2004: 119-120)

To be the object of a judgment about untrustworthiness or unreliability can be unpleasant or harmful. Yet we can make our distrust known without being charged with unfairness, so perhaps the same goes for expressions of resentment and indignation? If John is no longer trustworthy, he cannot object to Mary's distrust. Now suppose expressing blame simply consisted in *a judgment* that the wrongdoer manifested ill will. Consider the case of Phineas Gage. Gage survived an accident in which a tamping iron was driven through his prefrontal cortex. As a result, his personality changed. Before the accident he was trustworthy and hard-working; after the accident he became irreverent and anti-social. Now suppose that Gage hurt Mary by making an intentionally cruel joke. If Mary makes the judgment that Gage displayed ill will, Gage could not object to this claim. Nor does Mary's judgment appear to be unfair. Mary is making a factual claim about Gage's will, which happens to be true. But even though

the reactive attitudes might be responses to judgments about ill will, or have a propositional content concerning ill will, they are *not merely judgments*. Unlike distrust, they are essentially hostile feelings. When these feelings are expressed through reproaches, scolding or berating, it is a form of harsh treatment. Mary's judgment about Gage can be expressed without any hostile emotions. This judgment is not something Gage can object to. But if Mary were to express her reactive attitudes by scolding or berating him, Gage might plausibly object, given his accident, that it is unfair to treat him this way, or that he does not deserve this kind of treatment. So the analogy with distrust breaks down. In section 6 I will discuss one other way AE might account for the justification of expressed blame.

4. Blame as Sanctions

AE appears to fail the practicality requirement because it identified the relevant notion of blame with the feelings of the blamer. I will now consider a conception of blameworthiness that focuses on how blame is experienced by the wrongdoer. R. Jay Wallace argues that

[p]eople who are morally responsible may be made to answer for the actions, in the sense that their actions render them liable to certain kinds of distinctly moral responses. These responses include most saliently the response of moral blame [...] but they extend beyond simple blame to include a range of sanctioning responses as well, such as avoidance, reproach, scolding, denunciation, remonstrance and (at the limit) punishment. (1994: 54)

Similarly, Gary Watson takes blameworthiness "to involve the idea of liability to sanctions," by which he means "certain adverse or unwelcome treatment" (Watson 1996/2004: 275).¹¹

Because of this behavioural dimension of blame, Watson and Wallace argue, the norms

¹¹ Watson (1996/2004) famously distinguishes between two faces of responsibility: responsibility as accountability and responsibility as answerability. I will only be discussing his comments on responsibility in the accountability sense. Thanks to an anonymous reviewer for this clarification.

regulating blame should be understood in moral terms. Whether an agent is blameworthy is essentially a question about whether blaming is morally justified. The relevant norm, according to Wallace and Watson, is fairness. Inspired¹² by Watson and Wallace, we can formulate the following conception of blameworthiness:

FAIR SANCTIONS (FS): An agent S is blameworthy (for X) if and only if it would be fair to sanction S by outwardly expressing resentment or indignation towards S.

This conception has several strengths. First, it has obvious practical import. If we know that an agent is blameworthy, we also know that it is fair to sanction him. This does not mean that sanctioning behaviour is always the morally right way to respond to blameworthy agents. There might obviously be situations where there are counterweighing reasons. But the conception goes a long way to make sense of our blaming practices. FS also provides us with a straightforward argument for CONTROL. An agent is blameworthy to the extent that it is fair to sanction him for that action. But it is unfair to sanction someone for something she cannot control. So one cannot be blameworthy for something one cannot control.

This conception is, however, open to two objections. The first concerns the nature of blame. Let us distinguish between private blame and outwardly expressed blame. One might argue that private blame is the more fundamental notion, since it is possible to experience resentment and indignation without expressing these sentiments. The possibility of private blame suggests that the expression of reactive attitudes, in contrast to experiencing the attitudes, is not a necessary part of what blame is. Therefore, it may be argued, private blame

¹² It is not quite clear whether this is a formulation that Watson and Wallace would accept. They sometime emphasize that reactive attitudes come with a *disposition* to sanctioning behaviour. Watson claims that “blaming attitudes involve a readiness to adverse treatment” (1996/2004: 275). Wallace argues that blame “involves a disposition to engage in a variety of sanctioning activities” (1994: 94). This might very well be true as an empirical fact. But there is no necessary connection between blame and sanctions. Even though having the blame emotions involves a disposition to treat the wrongdoer in a certain way, this disposition need not be manifested (Graham 2014: 391; Nelkin 2013: 124). I therefore focus on the expression of the reactive attitude.

is the relevant notion, and fairness cannot be the right way of specifying the appropriateness condition. There might be ways of answering this objection. McKenna (2012: 69-70) argues that private blame is parasitic on its public manifestations: We explain the private instances by reference to their public manifestations. According to McKenna, a private episode of resentment is identified *as resentment*, rather than, say, contempt or disgust in terms of the public behaviour this emotion would give rise to.¹³ In what follows I will suppose for the sake of argument that some such justification can be given.¹⁴ In the next section, however, I will propose an account of blameworthiness that makes sense of the sanctioning aspect of blame without identifying blame with its outward expression, or insisting that its public manifestation is more fundamental.

Another worry concerns the notion of fairness. It seems possible to be blameworthy even when it would be unfair to express the reactive attitudes. Manuel Vargas (2004: 225) imagines a social policy according to which we never sanction people for their first moral offence. If we then arbitrarily suspend this practice and blame a random individual for his first offence, it would be unfair. But this does not mean that he was not blameworthy for his offence. We can also imagine cases where it would not be unfair to sanction blameless agents. Graham (2014: 391) notes that some agents are completely indifferent to other's disapproval. It would not be morally unfair to express anger and resentment towards them, even if they are not in the least blameworthy. The problem with using fairness is that it is sensitive to features that are not relevant to the issue of being blameworthy.

One natural way of avoiding this problem is by substituting "fairness" with "desert."¹⁵ I will follow Feinberg (1970), Scanlon (2013) and Pereboom (2013) in understanding desert

¹³ One might also *individuate* resentment, indignation, and so forth in terms of their common behavioural expressions without *identifying* these attitudes with the expressions. A functionalist account of them would do this. Thanks to an anonymous reviewer for pointing this out to me.

¹⁴ For a criticism of McKenna's account of blame, see Russel (forthcoming).

¹⁵ Wallace might not object to this. He suggests that "desert" might be one way of specifying what we mean by "fairness." (1994: 106-108)

as a relation of propriety that excludes consequentialist or contractualist considerations. According to Scanlon, “what is distinctive about a desert-based justification for treating a person a certain way is that it claims that treatment is made appropriate simply by certain facts about that person, or what he or she has done” (2013: 101).

Consider the examples above. Given the social policy, it may indeed be unfair to arbitrarily single out one person who is to blame for his first transgression. But that does not mean that he does not deserve to be blamed. Similarly, it is not unfair to express blame towards someone who is not blameworthy if he is completely indifferent or if he wants to be blamed. But if he is not blameworthy, he does not deserve to be blamed.

Can we say anything more about the notion of desert? I believe Pereboom (2013) is right when he claims that the concept of desert is *basic*. It cannot be reduced to or explained in terms of any more basic concepts. But, as McKenna (2012: 121) has pointed out, we should still say something about what desert entails. Following a suggestion from McKenna, I will say that desert entails noninstrumental goodness: If an agent deserves some harm, it will be noninstrumentally good that this harm occurs.

Substituting fairness with desert, we get:

DESERVED SANCTION (DS): An agent S is blameworthy (for X) if and only if S deserves to be sanctioned by outwardly expressed resentment or indignation (for X).¹⁶

The problem with FS was that the fairness of blaming was dependent on factors that are irrelevant to an agent’s blameworthiness. This is avoided by DS. But DS is open to another objection, which we can call *the problem of contingent harm*.¹⁷

¹⁶ DS is a slightly modified version of the account of blameworthiness found in Pereboom (2013: 189).

¹⁷ This problem will also affect FS.

The outward expression of resentment and indignation can be harmful. The severity of this harm, however, depends crucially on the psychological makeup of the recipient of blame, as well as the particular situation in which she is blamed. So on DS, the connection between sanctions and harm is only indirect and contingent. But, I will argue, what the agent deserves is not merely the sanction of expressed blame, but to suffer the harm *typically involved* in a sanction.

Suppose that John and Paul have committed crimes that are similar in every relevant way and that neither of them had any valid excuse for their actions. There is, however, one crucial difference. John cares about the judgments of others and is prone to feeling guilty. Paul, on the other hand, does not care much about what others think of him and is seldom bothered by guilt. Both, we stipulate, are blameworthy for their actions. So according to DS, both deserve to be the recipient of an outward expression of resentment or indignation. They would, however, experience this blame quite differently. John would be severely rattled and suffer from bouts of guilt. Paul would not mind the scolding and shrug the whole thing off. Do they both get what they deserve? It seems not. If John and Paul are equally blameworthy, they do not merely deserve the same treatment, they also deserve to suffer the same amount of harm from being blamed.

But why should this be the case? If blameworthy agents deserve to be the objects of outwardly expressed blame, isn't there a clear sense in which both John and Paul get what they deserve? After all, both are recipients of the reactive attitudes. The fact that people will experience such treatment differently, one might argue, is irrelevant to the issue of blameworthiness. I think this objection misses the gist of DS and FS. These conceptions focus on expressing, rather than feeling, resentment and indignation because the conditions of blameworthiness are taken to depend on how blame is experienced by the wrongdoer. It is the fact that the effect of blame is often harmful that makes it plausible to understand expressed

blame as a sanction that should be governed by fairness or desert. But it is easy to imagine circumstances where this picture of expressed blame does not apply. Consider a group of bullies who challenge each other to find the most effective way of humiliating their victim. When the victim expresses his resentment towards one of them after a particularly ingenious humiliation, the bully might not experience any harm at all, but rather pride as well as recognition from his peers.¹⁸ Although the bully is the recipient of an expression of resentment, he does not have the relevant experience of being blamed. What he deserves, I will suggest, is the harm of blame: blame *experienced* as a sanction.

Here is another way of illustrating the problem. Blameworthiness comes in degrees. Depending on the severity of the wrongdoing as well as the presence of mitigating factors, an agent can be more or less blameworthy. How would DS account for this? The natural answer would be to say that an agent is very blameworthy to the extent that she deserves to be the recipient of more intense feelings of resentment or indignation, or a harsher or more prolonged expression of these attitudes. But again, the relevant factor is how the expression of blame is experienced. Suppose that John committed a minor transgression, whereas Paul did something seriously wrong. Let us stipulate that Paul is more blameworthy than John. According to DS this would mean that Paul deserves harsher, more prolonged or more intense expressions of blame. But given his psychology, this would not amount to any more harm. John, on the other hand, also deserves to be the recipient of blame, albeit in a less severe form. But given his psychology, this would harm him much more than the more severe blame harmed Paul. So even if Paul were more blameworthy, John would suffer more harm as a result of the expressed blame.

The root of these problems is that the expression of blame is only contingently related to the harm of blame. Being the recipient of expressed blame can, and often will, be harmful.

¹⁸ See Nelkin (2013: 124) for a similar example.

But whether this is the case depends on matters that are not directly relevant to the agent's blameworthiness. The expression of resentment or indignation does not secure that the blameworthy agent gets what he deserves.

5. Emotions and Sanctions: Blameworthiness as Deserved Guilt

Let me sum up. So far I have discussed different conceptions of blameworthiness. According to AE, an agent is blameworthy to the extent that it would be appropriate to feel the reactive attitudes towards him. According to DS, an agent is blameworthy to the extent that the wrongdoer would deserve to be sanctioned by an expression of the reactive attitudes. I have argued that both these conceptions come with serious costs. If AE were correct, someone's being blameworthy would not explain the pro tanto permissibility of expressing blame. It therefore fails the practicality requirement. DS can account for this. But it cannot account for the fact that equally blameworthy agents deserve to suffer the same amount of harm. It is therefore vulnerable to the problem of contingent harm. I will now argue that it is possible to synthesize these two conceptions in a way that avoids these problems. This idea relies on the nature of guilt, and it is necessary to say a bit more about this emotion and its relation to the other reactive attitudes before I proceed.

I will follow Graham in understanding blame in terms of experienced reactive attitudes. To blame another person is to feel resentment or indignation towards him. To blame oneself is to feel guilt. These reactive attitudes are, I suggest, characterized by a combination of affect and some propositional content. What is this propositional content? One straightforward candidate would be the thought that the recipient of blame (oneself in the case of guilt) is blameworthy. This is not an option for the accounts of blameworthiness that I have been discussing in this paper. These accounts, as well as the one I will propose, seek to explain what it is to be blameworthy in terms of appropriate blame: an agent is blameworthy

because it is appropriate to blame him. If blame were partly constituted by the thought that the recipient of blame is blameworthy, the explanation of blameworthiness would be circular: an agent would be blameworthy because it is appropriate to think that he is blameworthy. Fortunately, there are several other plausible candidates: a violation of a moral requirement, the breach of an expectation that one holds other to (and oneself to in the case of guilt) (Wallace 1994: ch. 2), or the display of morally objectionable will (what Graham calls a violation of the requirement of respect). It may also be some combination of these, for example a violation of a moral requirement and an objectionable quality of will. I do not want to take a firm stance on this issue, but will suppose, with Graham, that the cognitive content of a blame emotion is that the wrongdoer displayed an objectionable quality of will (Graham 2014: 407).¹⁹ These blame emotions are thus unified by sharing the same propositional content.

Resentment, indignation and guilt are experienced from different perspectives, but nevertheless display certain symmetry. Resentment, indignation and guilt entail each other. Suppose that John hits Ben just as Sam is passing by. If it is appropriate for Ben to feel resentment towards John, it will be appropriate for Sam to feel indignation towards John. It will also be appropriate for John to feel guilt. The converse also holds. If it is appropriate for John to feel guilt, it is also appropriate for Ben to feel resentment and for Sam to feel indignation. This is a *pro tanto* propriety rather than an *all things considered* appropriateness. Suppose that Sam had just hit somebody. In that case it might be hypocritical, and thus not all things considered appropriate, for him to feel indignation towards John. But absent such defeaters, the mutual entailment holds. If I believe that it is appropriate for me to blame a wrongdoer, I must also think that it would be appropriate to blame myself if I did the same

¹⁹ Note that this cognitive content does not need to be identified with a *belief*. That would have made these accounts vulnerable to the problem of recalcitrant emotions: situations in which we feel guilty without believing we have done anything wrong or acted with ill will. Instead they may be understood as seemings. It can seem to you that you have acted wrongly, although you do not believe you acted wrongly. (Rosen 2015: 71).

thing (in the same circumstances). Similarly, if I think it appropriate to blame myself for a moral wrongdoing, I must also accept the pro tanto propriety of others blaming me.

There is, however, another respect in which there is a crucial, but often overlooked *asymmetry* between resentment and indignation on the one hand and guilt on the other. Indignation and resentment can be expressed to the wrongdoer as scolding, denouncement or berating. This will often be harmful to the wrongdoer. However, as we have seen, there is no *necessary* connection between the expression of blame and harm. Moreover, indignation and resentment need not be expressed. In these cases, the wrongdoer does not experience any harm at all. Guilt is different because it is directed towards oneself. Like resentment and indignation, guilt involves a propositional content—that I displayed ill will—as well as affect. This affect is unpleasant. As Clarke (2016: 124) has observed it is also inextricably linked to the propositional content.²⁰ When one experiences guilt one is pained by the thought that one has acted with ill will. The thought does not necessitate the unpleasant affect. It is possible to believe that one has acted with ill will, without being pained by this thought. But in that case it would not be guilt. In this sense, guilt is constituted by the painful affect. Unlike resentment and indignation, guilt is necessarily experienced by the wrongdoer. It is also intrinsically painful (Clarke 2013: 155). In contrast to resentment or indignation, the suffering involved in guilt is a necessary part of what it is to blame oneself. This suffering comes in different degrees. It can vary from a mild discomfort to a prolonged state of agony (Clarke 2013: 155). But if the emotional state does not involve suffering at all, it is not guilt.²¹ Let us now return to AE, which stated that:

²⁰ Clarke (2013, 2016) takes the constitutive thought of guilt to be that one is blameworthy. This thought explains the unpleasant affect.

²¹ Is it possible to *enjoy* the feeling of guilt, in a way parallel to how people might enjoy other intrinsically painful emotions such as fear or even sadness? I think it is possible to imagine a moral masochist who takes pleasure in the pain of guilt. This possibility would not undermine the claim that guilt necessarily involves suffering. The moral masochist would still suffer when he feels guilt. Unlike most people, however, he would enjoy this suffering. I thank an anonymous reviewer for this observation.

X is blameworthy for ϕ -ing iff^{def} the feeling of a blame emotion—resentment, indignation, or guilt—toward X in response to ϕ -ing would be appropriate. (Graham 2014: 390)

Given the asymmetry of suffering, how should we understand the appropriateness of the reactive attitudes? Let us start with guilt. Graham and Hieronymi took the relevant notion of appropriateness to be correctness. According to Graham, the reactive attitudes are constituted by the thought that the wrongdoer has violated a requirement of respect; the emotion is appropriate if this thought is true. This reasoning, however, relies on the idea that blame is not harmful to its recipient. But blaming oneself by feeling guilt is to suffer. This makes correctness a far less plausible candidate for the appropriateness of guilt. We cannot move directly from the truth of the constituent thought to the appropriateness of suffering: it is not clear that violating a requirement of respect makes it appropriate that the wrongdoer suffers. Because of the intimate connection between guilt and suffering, I think it is more plausible to understand the appropriateness of guilt in terms of desert: It is appropriate for a wrongdoer to feel guilt only if he or she *deserves* to feel guilt. This also makes better sense of the phenomenology of guilt. When I experience my feeling of guilt as appropriate, I do not merely acknowledge that I acted badly, or displayed ill will. It is a painful state to be in, and I take this *pain* to be appropriate: it is something I deserve.

The mutual entailment and the asymmetry of suffering present us with a puzzle. Resentment and indignation do not necessarily involve any suffering of the wrongdoer, but guilt does. This suggests that there might be different standards for blaming others than for blaming oneself. Because to feel guilt is to suffer, one might think that it would be more difficult to satisfy the conditions on self-blame. But this cannot be right: If it is appropriate

for me to blame others for action X in circumstances C, it must also be appropriate to blame myself for X in C.

I suggest that we resolve this puzzle by taking deserved guilt as the basic notion in our conception of blameworthiness. This suggestion is inspired by two recent papers by Randolph Clarke (2013; 2016). Clarke defends what he calls *the minimal desert thesis*: a blameworthy agent “deserves to feel guilty, to the right extent, at the right time, regarding the thing for which she is to blame, and deserves this because of her culpability.” (Clarke 2016: 122) According to Clarke, an agent deserves to feel guilty *because* he is blameworthy. Clarke also understands the cognitive content of feeling guilty to be the thought that I am blameworthy for something. These two features make it clear that while Clarke’s minimal desert thesis says something about what blameworthiness entails, it is not an attempt to develop an account of what blameworthiness is. Since the desert of guilt is explained by the fact that the agent is blameworthy, his account presupposes an independent grip on what blameworthiness is. Moreover, since the feeling of guilt is partly constituted by the thought that I am blameworthy, any analysis of blameworthiness which appeals to the feeling of guilt would be circular (Clarke 2016:122 see also Rosen 2015:80-81)

My proposal is to reverse Clarke’s direction of explanation, thereby using deserved guilt as the basis of an account of what it is to be blameworthy. Instead of starting with an independent notion of blameworthiness and arguing that those who are blameworthy deserve to feel guilt, I argue that we should start with the notion of deserved guilt, and use this notion to explain what it is to be blameworthy. To be blameworthy, I propose, just *is* to deserve guilt. On my account, the notion of deserved guilt is basic. When S is blameworthy for X, S is so *in virtue* of the fact that she deserves guilt for X. It is the notion of deserved guilt that explains

why the conditions of blameworthiness are as they are. The conditions of blameworthiness are as they are because for S to be blameworthy for X just is for S to deserve guilt for X.²²

By inserting deserved guilt into Wallace's schema we get a new conception of blameworthiness:

DESERVED GUILT (DG): An agent S is blameworthy (for X), and is an appropriate target of resentment and indignation, if and only if S deserves to feel guilt (for X).

An agent is blameworthy, I propose, because he deserves to feel guilt. Deserving guilt is both a necessary and sufficient condition on being blameworthy. If an agent is blameworthy, it will also be appropriate to feel resentment and indignation towards him. We can explain this by appealing to the mutual entailment of the reactive attitudes. If it is appropriate for others to blame me, it must be appropriate for me to blame myself: I must deserve to feel guilt. Conversely: if it is appropriate for me to blame myself, i.e. if I deserve guilt, it will also be appropriate for others to blame me.

DG provides an explanation of how blameworthiness relates *both* to emotions and to sanctions. Like AE, it understands blame as a matter of emotional reactions and not as an overt form of treatment. This allows for the possibility of private blame. Like FS and DS, it makes sense of the thought that being blameworthy involves deserved sanctions or suffering. But on DG, the suffering is of a special kind. When you feel guilt, you suffer *in recognizing what you have done*. I believe that this captures the specific kind of suffering that we take blameworthy agents to deserve. It is not suffering in general, but rather the specific pain of acknowledging your own objectionable quality of will. An agent may suffer as a result of being blamed, without recognizing his own fault, simply because he finds criticism unpleasant

²² Here I adopt Rosen's (2015) general structure of a theory of responsibility.

or because it affects his social standing. However, what he really deserves is the pain of guilt; the suffering involved in recognizing what he has done.²³

Another strength of DG is that it can give a straightforward explanation of excuses and exemptions. When people act wrongly without having the capacity to recognize moral reasons, or under duress, or while being coerced, they are not blameworthy *because* they do not deserve to feel guilt. This explanation is of course similar to what defenders of FS or DS would say. On these accounts, agents are excused or exempted because it would be unfair to sanction them, or because they would not deserve to be sanctioned. But on DG, we can apply this explanation without identifying blame with any form of overt treatment.

Note also how this explanation of exemptions and excuses is an improvement on AE. In section 3, I noted how an objectionable will is not a sufficient condition on appropriate blame. A young child or a demented adult can display ill will without being blameworthy. In order to account for these exemptions, AE would have to supplement the propositional content of the blame emotions with thoughts about moral competence or general capacities. This would, however, make the propositional content of blame emotions unduly complex. It is unlikely that anyone capable of feeling resentment also would be capable of having thoughts about moral competence or general capacities.²⁴ It would also appear to be something of a lucky coincidence that nature and/or culture equipped us with a set of emotions whose representational content just happened to match the condition under which someone deserves to suffer.²⁵ In comparison, on DG the representational content of the blame emotions is simple. Guilt involves the thought that I acted with ill will. Guilt is also painful. The

²³ Rosen (2015: 82-83) makes a related point: "The wrongdoer who responds to outward blame with a sincere and cheerful promise to do better next time but without a hint of guilt or remorse palpably frustrates a desire implicit in resentment"

²⁴ In developing an in some respects similar view, Gideon Rosen (2015: 75) points out that the propositional content of "resentment must be framed in terms that everyone capable of resentment understands." He calls this the naivety constraint. It is not easy to find a propositional content that can satisfy this constraint, and at the same time, make sense of all the commonly accepted exemptions and excuses.

²⁵ I owe this point to Knut Olav Skarsaune.

exemptions and excuses are explained by answering the moral question of when this painful emotion is deserved.²⁶

Finally, I would like to compare this explanation of exemptions and excuses to Clarke's minimal desert thesis. According to Clarke (2013; 2016), blameworthiness entails deserved guilt. He would therefore accept the biconditional which I used to define DG: An agent is blameworthy for X if and only if he deserves to feel guilty for X. This raises the question of whether the minimal desert thesis is equally suited to explain the conditions of blameworthiness. One might argue as follows: Given the minimal desert thesis, there have to be control conditions on blameworthiness because blameworthiness entails deserved guilt, and feeling guilty is intrinsically painful. This argument would, however, be problematic. According to the minimal desert thesis, agents deserve guilt *because* they are blameworthy. It therefore *presupposes* an independent account of what blameworthiness is and what its conditions are. These conditions will explain why blameworthy agents deserve to feel guilty. But there are several accounts of blameworthiness which reject control conditions. Let us for a moment assume that one of these is correct. Someone who does not believe in CONTROL could simply reverse the argument and *deny* that the blameworthy deserves to feel guilt: since blameworthiness, by stipulation, does not require control, the guilty does not deserve the suffering of guilt.²⁷ Of course, a proponent of the minimal desert thesis might *postulate* that

²⁶ A proponent of AE might reply by arguing that the content of the blame emotion is simply that the wrongdoer deserves to suffer for what she has done. Rosen (2015: 83) has recently made this suggestion. This thought would not be too complex and it would explain why the representational content of the blame emotion matches the conditions under which someone deserves to suffer. There are two problems with this suggestion. First, it is phenomenologically dubious. It is not obvious that one cannot blame another agent without thinking that she deserves to suffer. Second, it is too course-grained to explain the common excuses and exemptions. Whether an agent deserves to suffer would depend crucially on what kind of suffering we mean. One might make this representational constraint more specific, but this would run the danger of making it unduly complex.

²⁷ Clarke (2016: 134-135) considers this possibility. "Might it be said that although her thought that she is blameworthy is correct, the unpleasant affect that partly constitutes her feeling of guilt is inappropriate—she should not feel at all bad about the matter? I am skeptical that there is any kind of moral blameworthiness with respect to which this is so." Clarke does not elaborate on why he is skeptical about this. But earlier in the same paper he writes that poor quality of will: "does not suffice for one's being blameworthy, and a warranted thought with this content does not suffice to warrant a feeling of guilt." (Clarke 2016: 123). This suggests that Clarke finds accounts of moral blameworthiness which do not operate with at least some control conditions implausible on independent grounds. But it also suggests that if such accounts were true, he would be reluctant to the idea

there are control conditions on blameworthiness and argue that these conditions explains why blameworthy agents deserve to feel guilty. But this would not amount to an explanation of why there are control conditions on blameworthiness.

Alternatively, the proponent of the minimal desert thesis might *deny* that the absence of control conditions makes the suffering of guilt undeserved. On this view, the deserved pain of guilt is merely a fitting recognition of a true state of affairs, namely that the agent is blameworthy. Clarke (2016: 134) argues that there is no sense in which this recognition could be inappropriate, independently of what the conditions on blameworthiness are. This would be similar to Graham's claims about blame: appropriate blame is merely a recognition of true states of affairs: there is no sense in which it could be unfair. If successful, this line of argument would insulate the minimal desert thesis from the argument above. However, it would also mean that the pain of guilt would play no role in explaining or determining the conditions of blameworthiness. The pain of guilt would be a fitting recognition of the fact that the conditions of blameworthiness, *whatever they are*, have been breached.²⁸

DG, by contrast, is able to provide an explanation of why the conditions on blameworthiness are as they are. When agents are excused or exempted for their morally wrong behaviour it is because they do not deserve to feel guilt for this behaviour. Moreover, the pain of guilt plays a crucial explanatory role. It is because guilt is an intrinsically painful

that agents should feel the pain of guilt, merely because they acted with ill will.

²⁸ There is a third possibility. The cognitive content of guilt might not merely be the thin notion that I am blameworthy, but rather a more specific (and sophisticated) thought about control. If this could be established, I believe we would have an informative explanation of why there are control conditions on blameworthiness. However, this more specific cognitive content seems less plausible than Clarke's suggestion. First, this idea might seem to violate what Gideon Rosen (2015) calls the naivety constraint: the propositional content of guilt must be framed in terms that everyone capable of guilt understands. It seems possible for someone to feel guilt without understanding the notion of control. Second, people (across different times and cultures) tend to feel guilty for a wide variety of things. The thought of being blameworthy might be sufficiently thin to cover this variety, but the thought of being in control does not. Third, it is natural to assume that guilt has some biological bases, as well as being the product of cultural input. But neither biology nor culture gives us reason for confidence in this propositional content. It would be major coincident if the propositional content of guilt matches up perfectly with the correct control conditions. The alternative picture, on which we determine whether an agent is blameworthy by asking whether the painful feeling of realizing one's own wrongdoing is deserved, avoids this worry. In effect, it reduces the question of blameworthiness to the ethics of emotions. We have less reason to be afraid of debunking explanations of this kind of ethical reasoning.

emotion that we can expect the conditions of blameworthiness to be rather strict.

The explanatory role I ascribe to the pain of guilt has one implication that may seem problematic. I have suggested that the propositional content of guilt is that one acted with ill will. But the fact that one acts with ill will is not, on my view, sufficient for guilt to be *deserved*. It is therefore possible that an agent experiences guilt, has only true judgments, yet does not deserve guilt. How do we talk somebody out of an undeserved feeling of guilt if we cannot point to any mistake in reasoning or perception?²⁹ This difficulty would not arise on the minimal desert thesis. Clarke, as we have seen, takes the propositional content of guilt to be the thought that one is blameworthy. On this view, an agent who feels guilty without deserving guilt will have the mistaken thought that he is blameworthy.³⁰ However, I think there are a number of things a proponent of DG could say to an agent who feels guilt without deserving it. In fact we say things like this all the time: ‘Yes, you did X, but you shouldn’t feel bad about it’. An agent might be correct in assessing that he acted with ill will, but this, by itself, does not make it the case that he should be *pained* by this acknowledgement. Moreover, a person in the grip of an undeserved feeling of guilt will often be making a cognitive mistake, even according to DG. Feeling guilty is often accompanied by a belief that the pain of guilt is deserved and this belief can be mistaken.

6. Contingent Harm and the Practicality Requirement Revisited

²⁹ Thanks to an anonymous reviewer for asking me to address this implication of my account.

³⁰ Clarke (2016: 123) writes: “An appealing understanding of the warrant for a sentiment or emotion is epistemic warrant for its constitutive thought. (Compare: what warrants fear when one sees a tiger is the warranted thought that the animal is dangerous.) This consideration supports the suggestion I have offered.” This is an appealing understanding for most emotions (and as we have seen, this view of warrant motivates Graham’s account of blameworthiness). However, as I argued above, it is far less appealing when the emotion in question is also constituted by an intrinsically painful affect.

In this section I will show that DG also has the resources to answer the practicality objection and the problem of contingent harm. I will first turn to the problem of contingent harm. On DG, the connection between blame and suffering is necessary. Feeling guilty necessarily involves suffering. Now consider the objections to DESERVED SANCTION. In the first scenario, John and Paul were equally blameworthy. On DS, this means that they deserve the same form of sanction. But because of his psychological make-up, John suffers more than Paul from this sanction. This is problematic because, being *equally* blameworthy, they deserve the *same* amount of harm from being blamed. On DG, we can solve this problem. If John and Paul are equally blameworthy, they deserve to feel the same amount of guilt. This means that they also deserve to experience the same amount of suffering. In the second scenario, Paul was, by stipulation, more blameworthy than John. On DS, this means that Paul deserves a harsher sanction. But because of his psychological make-up, John would experience more harm from being blamed than Paul, even though Paul's sanction is more severe. Again, DG can solve this problem. If Paul is more blameworthy than John, he deserves to feel a greater amount of guilt. This means that he would also deserve a greater amount of suffering.

But why should the same amount of guilt equal the same amount of suffering, or more guilt equal more harm? This is because guilt is *constituted* by painful affect. One might perhaps argue that although harm is a necessary constituent of feeling guilt, feeling more guilt does not necessarily involve more suffering. But then it would be very difficult to understand what it means to feel *more* guilt. Frank may say that he experiences more guilt today for cheating on his taxes than he did yesterday. But if he adds that the guilt feels less painful today than it did yesterday, it would not be clear what he means by experiencing guilt. But perhaps two people could feel the same amount of guilt but experience different amounts of suffering? In one sense, this is of course possible. Feeling guilt might have different

consequences for different people. If John had an important job interview, his feeling of guilt might make him appear neurotic, and he might not get the job. This would be harmful to John. But what we are interested in is the suffering *directly involved* in feeling guilt. When we consider this kind of suffering, it seems clear that the same amount of guilt equals the same amount of suffering. Again, this is because guilt is constituted by painful affections. If John's feeling of guilt involved less painful affects than Paul's, they would not feel the same amount of guilt.

Next, consider the objection to APPROPRIATE EMOTIONS. The problem with this conception was that it failed the practicality requirement: if we know that an agent is blameworthy, we should also know that it is pro tanto permissible to express blame towards that agent. Can DG fulfil this requirement? I believe it can. But first we need to note a difficulty. On DG, what the agent deserves is not the expression of the reactive attitudes, but the feeling of guilt. Although it would be appropriate to resent the wrongdoer, DG itself says nothing about whether the wrongdoer deserves the *expression* of resentment. This means that when an agent is blameworthy, the expression of the reactive attitudes cannot be *directly* justified in terms of desert. I will now argue that there are two indirect ways DG can explain why expressing blame is pro tanto permissible.

The first explanation draws on the notion of desert. If a state is noninstrumentally good, other people will have a pro tanto reason to bring about that state. When wrongdoers deserve to feel guilty, other people will have a pro tanto reason to induce guilt in the wrongdoer (Clarke 2013: 158-159).

The second explanation draws on a standard notion of fairness. Because the agent deserves to feel guilty, it will normally be fair to express these attitudes. One way of finding out whether a particular treatment is fair is by asking whether the object of the treatment could reasonably object to it. So let us start with the perspective of the wrongdoer. Let us

assume that John is blameworthy for mistreating Alice: he deserves to feel guilt, and he is an appropriate target of resentment and indignation. He also considers himself to be blameworthy, so he concurs that he deserves to feel guilt because of what he has done. Suppose that Alice expresses her resentment towards him. Is this something John could plausibly object to? He already accepts that he deserves to feel guilt. Moreover, he also accepts that he is the appropriate target of resentment. Given this, it is hard to see how he could plausibly object when these attitudes are expressed. He might argue that he does not deserve *the particular suffering* he experiences when Alice expresses her resentment. But what would this objection consist in? It cannot be that the resentment and indignation lack justification. John already accepts that these emotions are justified. Nor can it be because it causes him to feel guilt, since he already accepts that he deserves to feel guilt. We can also consider the fairness of expressing attitudes from the viewpoint of the blamer. If Alice knows that John is blameworthy, she also knows that he is the appropriate target of her resentment. She knows that expressing her blame might cause suffering in John, because it can induce feelings of guilt. But feeling guilt is what John deserves. There are of course instances of expressed blame that blameworthy agents could object to. If the wrongdoer deserves a mild form of guilt, it can obviously be unfair to target him with prolonged and intense scolding. Moreover, not everyone has the standing to express blame; there are times when it would be hypocritical to do so. But in principle, it would be fair to express blame towards someone who deserves to feel guilt.

There are several ways of justifying expressed blame that do not satisfy the practicality requirement. We might express blame in order to influence people, protest, express our moral standing, or uphold a moral standard that has been breached.³¹ All of these justifications share a common feature: they do not rely on the blameworthiness of the agent.

³¹ See footnote 10 for a discussion of these views.

We may have reasons to influence people, confirm our moral standing, and so on, even if the wrongdoer was not blameworthy for his action. The justification that I have proposed differs in this respect. I have argued that it is fair to express the reactive attitudes *because* the agent is blameworthy. If the wrongdoer did not deserve to feel guilt, it would not be fair to express resentment and indignation towards him. I think this gives fairness the appropriate role in our theory of blameworthiness. Whether an agent is blameworthy is not determined by whether it is fair to express blame towards him. Fairness is sensitive to too many issues that are not relevant to the question of blameworthiness for it to play that role. But the question of whether it is fair to express blame surely *depends* on whether the agent is blameworthy. The fact that an agent is not blameworthy is a compelling reason for why it would not be fair to blame him.³²

But could not a proponent of AE simply adopt the same way of justifying expressed blame, and thereby satisfy the practicality requirement?³³ According to Graham, an agent is blameworthy if *a* blame emotion—resentment, indignation *or* guilt—would be appropriate. As I argued in section 3, the appropriate feeling of resentment or indignation does not, by itself, even make it pro tanto permissible to express these emotions. In order to account for the fairness of expressing these emotions, we first need to establish that *guilt is deserved*. It is because the wrongdoer deserves to feel guilt that he cannot reasonably object to the expression of resentment and indignation. According to Graham, it is appropriate for the wrongdoer to feel guilt if it is true that he violated a requirement of respect. But as I argued above: since guilt is an essentially painful state, correctness is not a plausible candidate for the

³² One might worry that this way of justifying expressed blame would open the door for the problem of contingent harm. Even if the expressed blame is proportional to the deserved guilt, equally blameworthy agents might, because of their different psychologies, suffer different amounts of harm as a result of expressed blame. But there is an important difference. On DG, the fairness of expressing blame plays no role in determining whether, or to what extent, an agent is blameworthy. An agent is blameworthy to the extent that she deserves to feel guilt. Equally blameworthy agents will deserve the same amount of guilt. So on DG, it will still be the case that equally blameworthy agents deserve to suffer the same amount of harm.

³³ Thanks to Jakob Elster and Torfinn Huvenes for pressing me to clarify this point.

appropriateness of guilt. A violation of a requirement of respect is not sufficient to establish that guilt is appropriate. Guilt is appropriate only if it is deserved. Graham cannot account for this. This also means that Graham cannot adopt my explanation of why it is fair to express blame towards the blameworthy. The harm involved in being the object of expressed blame is fair because the wrongdoer deserves guilt.

I have argued that the reactive attitudes mutually entail each other. But they are also asymmetrical with regards to suffering. Given this asymmetry, I have suggested that we should take deserved guilt as the basic notion in our conception of blameworthiness. DG, I have argued, can respond to *the problem of contingent harm* and has the resources to fulfil *the practicality requirement*.

Having presented and defended DG, I am now, finally, in a position to give a tentative argument for CONTROL. We are morally blameworthy for A only if A was (directly or indirectly) under our control because (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. The argument is tentative because I have not argued for the claim that one cannot deserve to suffer for something that is not within one's control. I take this general moral principle to have a strong intuitive appeal, but a full defence would require another paper. What I have done is to present a plausible conception of blameworthiness of which suffering is a necessary component. If DG is on the right track, the case for CONTROL will be considerably stronger.

7. Guilt: Fitting or Deserved?

There are other painful emotions than guilt. Grief, regret, embarrassment and shame are all, like guilt, intrinsically painful. Yet these emotions do not require a control condition in order to be appropriate. Suppose somebody spiked Jones' drink so that he got drunk, through no

fault of his own. While intoxicated he made an appalling racist remark.³⁴ It makes perfect sense for Jones to be embarrassed by what he did, regret what he did, feel shameful about his remark and so on. These emotions are fitting: his act is *embarrassment-worthy*, even though it was not under his control. So it is not true *in general* that the appropriateness of a painful emotion is subject to the control condition. Why should the appropriateness of guilt be governed by this condition? This is because the relevant sort of appropriateness of guilt is *desert*. It may be fitting for Jones to be embarrassed, but he does not deserve embarrassment.

Desert is sometimes understood in terms of fittingness (King 2012; Clarke 2013; 2016; Nelkin 2016).³⁵ If that were right, this response would not work, since it relies on a distinction between fittingness and desert. But I do not think it is right. Suppose your father just passed away. It is surely fitting to feel sadness or grief, but it would be very strange to say that you *deserved* these emotions. It might sometimes be fitting to feel shame, envy or regret, but we do not deserve these emotions. Fittingness is not a moral notion. (D'Arms and Jacobson 2000), but desert is. Above I suggested that we should understand desert as a basic notion that entails noninstrumental goodness. If Smith deserves to be punished it is noninstrumentally good that he is punished. Understood in this way it is clear that desert and fittingness can come apart.³⁶ It may be fitting to feel envy towards one's rival, but it is not noninstrumentally good; it can be fitting to feel amusement by a morally dubious joke, but not noninstrumentally good.³⁷

³⁴ I owe this example and the objection to Gideon Rosen.

³⁵ Clarke 2013 also suggests that desert can be understood in terms of noninstrumental goodness.

³⁶ Dana Nelkin (2016: 178) makes the opposite argument. She takes desert to be a special kind of fittingness. Because fittingness in general does not entail noninstrumental goodness, she rejects the claim that desert entails noninstrumental goodness. She uses the example of a racist joke. It may be fitting to laugh of it, although it is not in any way good. Similarly, she argues against the claim that "X being deserving of sanction provides a reason to sanction." A spontaneous smile or an expression of surprise may be fitting, but does not provide reasons. However, none of these examples seems to be instances of *desert*. While it may be fitting to laugh of a racist joke, smile when seeing a child for the first time after a long trip, or show an expression of surprise in response to a carefully planned surprise party, it would be odd to say that one *deserves* any of these reactions. I take this to indicate that desert and fittingness are different notions of appropriateness.

³⁷ To argue that a amusement is not fitting because it would be immoral is an example of what D'Arms and Jacobson (2000) call "the moralistic fallacy"

This reply raises another question: Can guilt be fitting but undeserved? If that were the case we might have two competing accounts of blameworthiness, one according to which X is blameworthy for A iff guilt is fitting; another according to which X is blameworthy iff guilt is deserved. The second notion would be governed by a control condition for the reasons I have given; the first would not. And so we can imagine opponents of the control condition responding to my argument as follows: “You’re quite right about blameworthiness of the second sort: it’s governed by a control condition, and no one is non-derivatively blameworthy in that sense for attitudes, thoughtless omissions, and the rest. But that doesn’t show that we can’t be blameworthy in the first sense for things we don’t control, and that’s all we ever meant to say.”

While I cannot exclude the possibility of this competing conception, I think there are reasons to favour DESERVED GUILT. First, desert allows for an informative explanation of why the conditions of blameworthiness are as they are. As Gideon Rosen has noted, the fittingness relation cannot help us to *explain and articulate* these conditions. There “is no hope for an articulate account of the conditions under which (say) hilarity is a fitting response to a joke, much less for an account of why hilarity is fitting under those conditions.” (Rosen 2015: 71). Similarly, there would be no way of articulating and explain the conditions under which guilt is fitting. DESERVED GUILT, by contrast, explains the conditions of blameworthiness by appealing to norms for when the suffering of guilt is deserved, which in turn may be explained by reference to general moral norms of deserved suffering.

Second, it seems to me that guilt (to the right degree, in the right circumstances) will always be noninstrumentally good in a way that goes beyond the propriety of a fitting attitude. Clarke (2013: 155-157) suggests two reasons for why guilt is noninstrumentally good. First, feeling guilty is to acknowledge one’s own moral fault, or objectionable quality of will.³⁸ This

³⁸ Note that Clarke (2013, 2016) takes the thought included in the emotion of guilt to be that one is blameworthy, not that one acted with ill will.

involves recognizing a morally important fact about one self and there is value in this recognition. Second, the painful affect that is constitutive of feeling guilt, expresses a valuable moral concern. Both the recognition and the expression of concern appear to be noninstrumentally good in a way the propriety of a fitting attitude is not.

Third, desert gives other people a pro tanto reason to bring about the noninstrumentally good state. This is not generally the case for fittingness. Shame, embarrassment and regret are painful emotions that can be fitting. But the fact that embarrassment would be a fitting response to an action over which you had no control, does not give other people any reason to induce embarrassment in you. If I did something shameful or regrettable this would not provide other people with even a pro tanto reason to induce these emotions in me. Guilt differs from other painful emotions in this respect. If a blameworthy agent does not feel guilt, other agents do have a pro tanto reason to bring it about that he experiences guilt. The standard way of doing this, of course, is by expressing blame.³⁹

Grief might be more similar to guilt than embarrassment, shame and regret.⁴⁰ Consider the grief one feels when a loved one dies. This emotion is fitting, but one might also argue that it is noninstrumentally good. If I compare two worlds, one in which children grieve their deceased parents and one in which they do not, the one in which they grieve may seem to be noninstrumentally better. Just as guilt can be a noninstrumentally good recognition of one's fault in doing wrong, it seems plausible, as Dan Moller (2007) has argued, that grief can be a valuable recognition of a great good that is irretrievably lost. Both emotions are intrinsically painful, and this pain is a part of what makes them appropriate. Still, I think there are important differences between guilt and grief. First, guilt is deserved in virtue of the agent's

³⁹ Miranda Fricker (2016) has recently developed an account of blame, according to which the illocutionary point of blame is to inspire remorse: "it aims to bring the wrongdoer to see or fully acknowledge the moral significance of what they have done or failed to do" (Fricker 2016: 173). For similar claims about the purpose or function of blame, see Rosen (2015: 82-83) and Shoemaker (2015: 110-111).

⁴⁰ I am grateful to an anonymous reviewer for asking me to elaborate on the differences between guilt and grief.

action or omission. Grief, on the other hand, is fitting simply in virtue of the loss of a loved one. This points to a broader difference between fittingness and desert. It seems plausible that, to the extent that an agent can deserve to suffer, it can only be for something he has done or omitted. *Deserved* suffering, at a minimum, requires an action or omission as its desert base. *Fitting* suffering makes no such deontic restrictions. Second, let us consider whether other people have a reason to induce grief in a person who has lost a loved one. Initially, this may seem implausible. Yet if we accept that grief is noninstrumentally good, it may not be obvious why other people could not have a pro tanto reason to do so. However, there is still a crucial difference between guilt and grief. Suppose that I am aware that grief is a valuable recognition of a great good that is lost, but also believe that the emotion is too painful. In that case, other people will not have any reason to induce the feeling of grief in me. A parallel argument for guilt would not work. Whether I *want* to feel the pain of guilt is irrelevant to the reasons other people have for inducing guilt in me. This might be another way of bringing out the distinctively moral character of desert.⁴¹ Grief can be fitting but not deserved. Whereas a world in which children grieve their deceased parents might be better than a world in which they do not, it is not clear that this world is *morally* better. A world in which agents who deserve guilt feel guilty, on the other hand, is a morally better world. Desert, I suggest, entails a distinctly moral kind of noninstrumental goodness, which provides other people with a pro tanto reason to induce the noninstrumentally good state in the wrongdoer, irrespectively of whether the wrongdoer wishes to experience this state or not.

8. Conclusion

I have presented a new conception of blameworthiness. To be blameworthy, I have argued, is most fundamentally to deserve to feel guilty. On this conception, the relevant notion of blame

⁴¹ Clarke (2016: 127), following Feinberg (1970) specifies the moral propriety of desert, by calling it a "consideration of justice"

is understood as an emotion, not as any form of overt treatment. Yet, this conception also explains why blameworthiness necessarily involves suffering. The account has the resources to solve the problem of contingent harm and it fulfils the practicality requirement. I have argued that it supports an argument for CONTROL. Although the implications of the account, as well as the relevant notion of control still need to be worked out, I hope to have shown that this conception provides us with a *framework* for addressing these issues: The conditions of blameworthiness are the conditions for deserving guilt.

Acknowledgements

Versions of this paper was presented at the Nordic Network for Political Philosophy and at the PPPE Club, Centre for Mind and Nature, University of Oslo. I would like to thank the audiences on each of these occasions for helpful discussion. For written comments or discussion I am very grateful to Marcia Baron, Einar Duenger Bøhn, Lars Christie, Anna Drozdowicz, Jakob Elster, Christel Fricke, Robert Huseby, Torfinn Huvenes, Neil Levy, Jon Anstein Olsen, Gideon Rosen, Knut Olav Skarsaune, and two anonymous reviewers for this journal.

References

Arpaly, Nomy. 2003. *Unprincipled Virtue. An Inquiry into Moral Agency*. Oxford: Oxford University Press.

Brink, David and Dana Nelkin. 2013. "Fairness and Architecture of Responsibility." In Shoemaker, D. (ed.) *Oxford Studies in Agency and Responsibility* vol. 1: pp. 284-313.

Clarke, Randolph. 2013. "Some Thesis on Desert". *Philosophical Explorations*, 16 (2): 153-64.

- Clarke, Randolph. 2016. Moral Responsibility, Guilt and Retributivism. *The Journal of Ethics* 20 (1): 121-137
- D'Arms, Justin and Daniel Jacobson. 2000. "The Moralistic Fallacy" In *Philosophy and Phenomenological Research* Vol. 61, nr 1, 66 - 90.
- Feinberg, Joel. 1970. *Doing and Deserving*. Princeton N. J.: Princeton University Press.
- Fischer, John Martin and Mark Ravizza. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge University Press.
- Fricker, Miranda. 2016. What's the Point of Blame? A Paradigm Based Explanation. *Nous* 50 (1): 165-183
- Graham, Peter. 2014. "A Sketch of a Theory of Blameworthiness". *Philosophy and Phenomenological Research* 88 (2):388-409.
- Hieronymi, Pamela. 2004. "The Force and Fairness of Blame". *Philosophical Perspectives*, 18: 115-148.
- King, Matt. 2012. Moral Responsibility and Merit. *Journal of Ethics and Social Philosophy* 6 (2): 1-17.
- McKenna, Michael. 2012. *Conversation and Responsibility*. Oxford: Oxford University Press.
- Moller, Dan. Love and Death. *Journal of Philosophy* 104 (6): 301-316.
- Nelkin, Dana. 2011. *Making Sense of Freedom and Responsibility*. Oxford: Oxford University Press
- Nelkin, Dana. 2013. "Desert, Fairness and Resentment". *Philosophical Explorations* 16 (2): 1-16.

- Nelkin, Dana. 2016. "Accountability and Desert". *The Journal of Ethics* 20 (1): 173-189.
- Pereboom, Derk. 2007. "Hard Incompatibilism" in Pereboom, D., J.M Fischer, R. Kane, and M.Vargas (2007). *Four Views on Free Will*. New York: Blackwell, pp 85-126.
- Pereboom, Derk. 2013. "Free Will Skepticism, Blame, and Obligation". In D. J. Coates and N. A. Tognazzini, (eds.), *Blame: Its Nature and Norms*, New York: Oxford University Press, pp. 89-206.
- Pereboom, Derk. 2014. *Free Will, Agency, and Meaning in Life*. Oxford: Oxford University Press.
- Rosen, Gideon. 2002. "Culpability and Ignorance". *Proceedings of the Aristotelian Society*, 103: 61-84.
- Rosen, Gideon. 2015. "The Athletic Conception of Moral Responsibility". In Clarke, Randolph, Michael McKenna and Angela Smith (eds.), *The Nature of Moral Responsibility. New Essays*, Oxford: Oxford University Press, pp. 65-89.
- Russel, Paul. Forthcoming. "Can we Talk? *Philosophical Review*.
- Scanlon, Thomas. M. 2008. *Moral Dimensions. Meaning, Permissibility and Blame*. Harvard University Press.
- Scanlon, Thomas. M. 2013. "Giving Desert its Due". *Philosophical Explorations* 16 (2): 101-116.
- Strawson, Peter. 1962/1993. "Freedom and resentment". In J. M. Fischer & M. Ravizza (Eds.), *Perspectives on Moral Responsibility*. Ithaca: Cornell University Press, pp. 45-66
- Sher, George. 2006. *In Praise of Blame*. Oxford: Oxford University Press.

- Sher, George. 2009. *Who Knew? Responsibility Without Awareness*. Oxford: Oxford University Press.
- Shoemaker, D. 2015. *Responsibility from the Margins*. Oxford: Oxford University Press.
- Smith, Angela. 2005. "Responsibility for Attitudes: Activity and Passivity in Mental Life". *Ethics* 115: 236-271.
- Smith, Angela. 2007. "On Being Responsible and Holding and Responsible." In *Journal of Ethics* 11: 456-84.
- Talbert, Matthew. 2012. "Moral Competence, Moral Blame, and Protest". *Journal of Ethics* 16 (1): 89-109.
- Vargas, Manuel. 2004. "Responsibility and the Aims of Theory: Strawson and Revisionism". *Pacific Philosophical Quarterly* 85 (2): 218-241.
- Wallace, R. Jay. 1994. *Responsibility and the Moral Sentiments*. Harvard University Press.
- Wallace, R Jay. 2011. "Dispassionate Opprobrium". In Wallace R., J., R. Kumar & S., R. Freeman (eds.). *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*. Oxford University Press, pp. 348-372.
- Watson, Gary. 1996/2004. "Two Faces of Responsibility." In Watson (2004) *Agency and Answerability* New York: Oxford University Press, pp. 260-289.
- Wolf, Susan. 2011. "Blame Italian Style." In Wallace R., J., R. Kumar & S., R. Freeman (eds.). *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*. Oxford University Press, pp. 332-347.
- Zimmerman, Michael., J. 1997. "Moral Responsibility and Ignorance". *Ethics* 107 (3):410-426

