

# Accountability, Answerability and Freedom

---

**Published in Social Theory and Practice 42(4): 681-705, 2016**

## ABSTRACT

It has been argued that we cannot be morally responsible in the sense required to deserve blame or punishment if the world is deterministic, but still morally responsible in the sense of being apt targets for moral criticism. Desert-entailing moral responsibility is supposed to be more freedom demanding than other kinds of responsibility, since it justifies subjecting people to blame and punishments, is non-consequentialist, and shown by thought experiments to be incompatible with determinism. In this paper, I will show that all these arguments can be resisted.

*Keywords: moral responsibility, desert, blame, punishment, moral criticism.*

# Accountability, Answerability and Freedom

Published in *Social Theory and Practice* 42(4): 681-705 (2016)

## 1. Introduction

A number of philosophers distinguish between different ways of understanding moral responsibility, or between different kinds of moral responsibility. J. J. C. Smart claims that we can understand the practice of holding responsible in a clearheaded and pragmatic way or a metaphysically confused way.<sup>1</sup> Gary Watson argues that responsibility has “two faces”, one of aretaic evaluation and one of accountability.<sup>2</sup> Robert Kane writes that ultimate moral responsibility requires libertarian free will, and is different from compatibilist deconstructions of responsibility.<sup>3</sup> Thomas Scanlon discusses the difference between blame as mere character assessment, blame as a response to behaviour that changes one’s relationship with the person blamed, and blame as a social sanction.<sup>4</sup> David Shoemaker argues that our actual moral responsibility practice embodies three distinct conceptions: attributability, answerability and accountability.<sup>5</sup> Derk Pereboom distinguishes moral answerability from moral responsibility in the basic desert sense.<sup>6</sup>

The debate suffers from a lack of consensus when it comes to terminology, and it is often unclear whether two philosophers discuss the same thing. However, a fairly wide-spread

---

<sup>1</sup> J. J. C. Smart, “Freewill, Praise and Blame,” *Mind* 70 (1961): 291-306

<sup>2</sup> Gary Watson, “Two Faces of Responsibility,” *Philosophical Topics*. 24 (1996): 227-248.

<sup>3</sup> Robert Kane, *The Significance of Free Will* (New York: Oxford University Press, 1998).

<sup>4</sup> Thomas M. Scanlon, *Moral Dimensions: Permissibility, Meaning, Blame* (Cambridge: Belknap Press of Harvard University Press, 2008)

<sup>5</sup> David Shoemaker, “Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility,” *Ethics*. 121 (2011): 602-632.

<sup>6</sup> Derk Pereboom, “Free Will Skepticism and Criminal Punishment,” in Thomas Nadelhoffer (Ed.) *the Future of Punishment* (New York: Oxford University Press, 2013), pp. 49-78, and *Free Will, Agency and Meaning in Life* (New York: Oxford University Press, 2014), Ch. 6.

view seems to be that we can meaningfully discuss at least two different kinds of moral responsibility, one of which is more freedom demanding than the other: on the one hand, desert-entailing moral responsibility, and on the other, a kind of responsibility concerned with whether the agent acted on good reasons and lived up to certain justified moral demands.<sup>7</sup> There are a number of arguments for the thesis that desert-entailing moral responsibility is more freedom demanding than the latter kind. My aim in this paper is to show that they can all be resisted.

Following Shoemaker and Watson, I will consistently use the term *accountability* for the first kind of responsibility, and following Shoemaker and Pereboom, *answerability* for the second kind.<sup>8</sup> *Desert-entailing moral responsibility* or simply *moral responsibility* are more common than *accountability* in the philosophical literature, but *accountability* has the distinct advantage of being both short and acknowledging that there can be other kinds of moral responsibility.<sup>9</sup> I will provide more precise definitions of these terms in the next section, before engaging the arguments for the thesis that accountability is more freedom demanding than answerability.

---

<sup>7</sup> See, e.g.: Derk Pereboom, *Living Without Free Will*, (Cambridge: Cambridge University Press, 2001), pp. 139-140, 157; Pereboom, “Free Will Skepticism and Criminal Punishment” p. 51; Pereboom, *Free Will, Agency and Meaning in Life*, Ch. 6; Bruce Waller, “Virtue Unrewarded: Morality Without Moral Responsibility”, *Philosophia* 31 (2004): 427-447; Waller, “Sincere Apology Without Moral Responsibility”, *Social Theory and Practice* 33 (2007): 441-465; Shoemaker, “Attributability, Answerability and Accountability: Toward a Wider Theory of Moral Responsibility”; Derek Parfit, *On What Matters. Volume One* (Oxford: Oxford University Press, 2011) Ch. 11; Derk Pereboom and Gunnar Björnsson, “Traditional and Experimental Approaches to Free Will and Moral Responsibility”, forthcoming in Justin Sytsma and Wesley Buckwalter (Eds.) *Companion to Experimental Philosophy* (Blackwell).

<sup>8</sup> Shoemaker, “Attributability, Answerability and Accountability: Toward a Wider Theory of Moral Responsibility”; Watson, “Two Faces of Responsibility”; Pereboom, “Free Will Skepticism and Criminal Punishment”.

<sup>9</sup> Since there is no universally adopted terminology, it should be noted that other philosophers might use these terms differently than I do.

## 2. Accountability and answerability

An agent is *accountable* for her action when she deserves to be praised or blamed, rewarded or punished for what she did (or, if the action was morally neutral, she stands in such a relation to her action that she *would* have deserved praise or blame if the action had had a different moral status). She deserves this treatment in a pre-institutional and backward looking sense of ‘deserve’ – it is not merely the case that she ought to receive a certain treatment according to previously established laws or rules, and neither is it the case that treating her in this way must be justified by reference to beneficial consequences. A number of philosophers argue that an agent cannot deserve praise or blame, punishment or reward, unless she has a libertarian, or perhaps an impossible, kind of free will.<sup>10</sup>

An agent is *answerable* for her action when it makes sense to ask her why she did what she did, criticize her reasons if they were bad ones, demand that she does not repeat this kind of behaviour and so on. In short, when an agent is answerable for an action, she is an apt target for moral criticism. We morally criticize in order to achieve three important goals; we want to *morally improve* the wrongdoer (and ourselves; moral criticism might give the criticizer important moral insights too), prevent future wrongdoing and restore or improve relationships.<sup>11</sup> Philosophers tend to focus on answerability for wrongdoing, but we can easily think of a positive analogue. We might say that an agent is answerable for an exemplary act when she is an apt target for *moral commendation*. We can morally commend someone

---

<sup>10</sup> Pereboom, *Living Without Free Will*; “Free Will Skepticism and Criminal Punishment”; *Free Will, Agency and Meaning in Life*, Ch. 6; Waller, “Virtue Unrewarded: Morality Without Moral Responsibility”; “Sincere Apology Without Moral Responsibility”; Derek Parfit, *On What Matters. Volume One* (Oxford: Oxford University Press, 2011) Ch. 11; Galen Strawson, “The Bounds of Freedom”, in Robert Kane (Ed.) *The Oxford Handbook of Free Will: First Edition* (New York: Oxford University Press, 2002), pp. 441-460.

<sup>11</sup> Waller, “Sincere Apology Without Moral Responsibility”, p. 456; Erin Kelly, “Doing Without Desert”, *Pacific Philosophical Quarterly* 83 (2002): 180-205 p. 194; Pereboom, *Living Without Free Will*, p. 156; “Free Will Skepticism and Criminal Punishment”, p. 51; *Free Will, Agency and Meaning in Life*, pp. 132-134; Parfit, *On What Matters*, p. 261; Pereboom and Björnsson, forthcoming.

because we want to encourage her to perform more exemplary acts in the future, and because we want to further strengthen our relationship with her.

When philosophers discuss moral improvement in this context, it is not conceived of as a mere decreased frequency of undesirable behaviours and an increased frequency of desirable ones. Rather, moral improvement is seen as a coming to understand moral reasons better and changing one's behaviour in light of these reasons, thus becoming a better person. Likewise, the prevention of future wrongdoing and restoration and improvement of relationships are supposed to come about because the wrongdoer comes to understand the moral reasons against doing what she did, and change her behaviour in light of those reasons. Therefore, an agent must be responsive to reasons (and perhaps not just reasons in general, but *moral* reasons) in order to be answerable for what she does. She must have the general ability to understand moral criticism, and to regulate her behaviour accordingly. Pereboom argues that reason responsiveness is not only necessary but sufficient for answerability; as soon as the agent can respond to reasons, it makes sense to morally criticize her wrongdoing with the aim of morally improving her, preventing future wrongdoing and improving or restoring relationships. He points out that no similarly obvious connection exists between reason responsiveness and accountability.<sup>12</sup> I agree. Some compatibilists argue that reason responsiveness suffices for accountability as well,<sup>13</sup> but these arguments rely mostly on intuitions that not everyone shares. However, my purpose in this paper is not to provide a positive argument for compatibilism by establishing a conceptual link between accountability and reason responsiveness; I will merely show that no argument for accountability being more freedom demanding than answerability succeeds.

---

<sup>12</sup> Pereboom, *Free Will, Agency and Meaning in Life*, p. 136.

<sup>13</sup> E.g., John M. Fischer and Mark Ravizza, *Responsibility and Control. A Theory of Moral Responsibility* (New York: Cambridge University Press, 1998).

So far, I have written about blame *or praise* and punishment *or reward*. However, from now on, I will focus exclusively on wrongful actions, blame and moral criticism rather than their positive analogues. I do so in order to sidestep the debate about whether blame is more freedom demanding than praise.<sup>14</sup> I will consistently use the term *blame* to refer to the practice of holding people *accountable* for wrongful actions and the term *moral criticism* to refer to the practice of holding them *answerable*, since I believe this will make the text easier to follow. It should be noted that I here depart from Pereboom, who uses the term ‘blame’ for both.<sup>15</sup>

I provided, in the previous section, a list of philosophers who distinguish accountability from answerability, but there are also philosophers who conflate the two. Accountability justifies blame for wrongful actions, whereas answerability justifies moral criticism – but if blame is the same thing as moral criticism, one might conclude that accountability is the same thing as answerability as well, and thus, trivially, has the same freedom requirements.<sup>16</sup> There are undoubtedly clear similarities between blame and moral criticism – both refer to the moral reasons that the person blamed or criticized supposedly had against doing what she did. We blame people by saying things like “That was cruel!” “You could have killed her!” or “You promised not to do that!” All these statements cite moral reasons – aretaic, consequentialist and deontological respectively – as to why the behaviour that incurred blame was wrong. I will, however, grant my opponents that blame is different from moral criticism, and that, since

---

<sup>14</sup> E.g., Susan Wolf, *Freedom Within Reason* (New York: Oxford University Press, 1990), and Dana K. Nelkin, *Making Sense of Freedom and Responsibility* (New York: Oxford University Press, 2011).

<sup>15</sup> Pereboom, *Free Will, Agency and Meaning in Life*.

<sup>16</sup> Antony Duff, *Trials and Punishment* (Cambridge: Cambridge University Press, 1986), pp. 40-41; Ronald Dworkin, *Justice for Hedgehogs* (London: The Belknap Press of Harvard University, 2011), pp. 224-225; Angela M. Smith, “Control, Responsibility and Moral Assessment”, *Philosophical Studies* 138 (2008): 367-392; “Attributability, Answerability and Accountability: In Defense of a Unified Account”, *Ethics*, 122 (2012): 575-589.

accountability justifies blame whereas answerability justifies moral criticism, accountability is different from answerability as well. Firstly, blame is accompanied by certain attitudes such as resentment, indignation and moral anger.<sup>17</sup> Moral criticism, as defined by philosophers who distinguish accountability from answerability, need not involve an expression of these attitudes. Secondly, we have already established that accountability, and thus blame, are backward looking. If blame can be justified at all, it can be justified without appeal to beneficial consequences. Moral criticism, on the other hand, is justified by the three goals of moral improvement, prevention of future wrongdoing and restoration or improvement of relationships.

At this point, *accountability*, *answerability*, *blame* and *moral criticism* have been sufficiently well defined to allow for a discussion of the arguments for the thesis that accountability is more freedom demanding than answerability. There are four important arguments for this thesis in the philosophical literature. They are *the argument from the intentional infliction of suffering*, *the argument from harsh punishments*, *the argument from a lack of consequentialist justification* and finally *the argument from manipulation cases*. In the following sections I will deal with them one by one, and show that they can all be resisted.

### **3. The argument from the intentional infliction of suffering**

Agents who are accountable for what they do deserve to be blamed when they do wrong. When we explicitly blame someone, we express certain reactive attitudes that tend to be painful for the blamee; we try to make her *feel bad* about what she did. When an agent deserves to be blamed, she thus deserves to suffer; perhaps only psychologically, but

---

<sup>17</sup> Pereboom, *Living Without Free Will*, p. 208; *Free Will, Agency and Meaning in Life*, p. 128; Peter F. Strawson, "Freedom and Resentment", reprinted in Paul Russell and Oisín Deery (Eds.) *The Philosophy of Free Will: Essential Readings from the Contemporary Debate* (New York: Oxford University Press, 1962/2013), pp. 63-83.

psychological suffering is still suffering. It has been argued that making wrongdoers suffer can only be justified if what they did was, in a very strong sense, up to them, and that accountability is therefore more freedom demanding than answerability.<sup>18</sup>

We may, of course, silently blame people in our heads or blame the already dead, in which case we do not try to make the blamees feel bad. Plausibly, though, these kinds of blame are parasitic on the paradigm case of explicit blame. Furthermore, even silent blame or blame of the dead plausibly involves a preference for the (past, in the case of the dead) suffering of the blamee. And insofar as it is at all possible for the blamee to feel bad, we try to make her feel that way when blaming her.

All of the above is controversial. There are philosophers who analyse blame differently.<sup>19</sup> If blame need not involve an attempt to make the blamee feel bad, this argument for accountability being more freedom demanding than answerability fails. However, I do not believe that alternative analyses of blame are relevant in this context. ‘Blame’ in a wide sense may include what I in this paper refer to as ‘moral criticism’, and perhaps other kinds of reactions and responses to wrongdoing as well. Here, I am narrowly concerned with accountability, and the kind of response that an accountable wrongdoer deserves to receive. I find it plausible, and will in any case grant my opponents for the sake of discussion, that an accountable wrongdoer deserves to be blamed in this more narrow, pain-intending sense. What I will argue is that moral criticism, too, involves an attempt to make the person criticized suffer, and that the intended painfulness of blame therefore does not support the thesis that accountability is more freedom demanding than answerability.

---

<sup>18</sup> Parfit, *On What Matters: Volume One*, pp. 264-272; Pereboom, *Living Without Free Will*, pp. 208-209; Waller, “Sincere Apology Without Moral Responsibility”. Furthermore suggested by Scanlon, *Moral Dimensions: Permissibility, Meaning, Blame*, p. 183 and Watson, “Two Faces of Responsibility”, pp. 238-239.

<sup>19</sup> See, for instance, Scanlon, *Moral Dimensions: Permissibility, Meaning, Blame* and Pamela Hieronymi, “The Force and Fairness of Blame”, *Ethics* 18 (2004): 115-148.

Waller and Pereboom, who both believe that answerability exists whereas accountability would require an unlikely or even impossible kind of free will, do admit that the realization that one has done wrong is a painful one.<sup>20</sup> As we have seen, moral criticism aims to make the agent understand the moral reasons against the action she performed; to realize that she did wrong, and change her behaviour in light of this realization. Since this is painful, moral criticism too attempts to make the person criticized suffer. Answerability, not just accountability, justifies the intentional infliction of suffering.

Pereboom writes that we could only be justified in showing *moral anger* or *moral resentment* towards other agents if they were (counterfactually) accountable for what they did, but we can still be justified in expressing alarm, distress and sadness.<sup>21</sup> However, most of us know from experience that being subjected to another's sadness over what we have done may be at least as painful as being subjected to her anger. Since these experiences can be equally painful, why would the latter require more freedom on part of the person criticized for its justification than the former? One might argue that we *intend* to make others feel bad when exposing them to our anger and resentment, whereas we merely *foresee* that they will feel this way when we express our sadness, alarm or distress, or morally criticize them without expressing any particular emotion. What we *intend*, when we morally criticize, is for the agent to morally improve, abstain from harming others in the future and for our relationships to improve. However, I do not believe that this line of argument is tenable; the pain we cause

---

<sup>20</sup> Pereboom, *Living Without Free Will*, p. 205; *Free Will, Agency and Meaning in Life*, pp. 134, 186; Waller, "Sincere Apology Without Moral Responsibility", pp. 448-451.

<sup>21</sup> Pereboom, *Living Without Free Will*, p. 208; *Free Will, Agency and Meaning in Life*, p. 134.

people by expressing sadness or criticizing them is simply *too close* to the intended result of moral improvement to plausibly count as foreseen but not intended.<sup>22</sup>

It might be possible to make a conceptual distinction between an agent's fully realizing that she did wrong and the psychological pain that she experiences as a result (although it has been argued that psychological suffering partly *constitutes* a full moral realization of the wrongness of one's previous actions<sup>23</sup>). It may even be possible for certain agents with particular psychologies to fully realize the wrongness of their conduct and sincerely resolve not to do what they did again completely without suffering. It is still implausible to claim that I did not intend, but merely foresaw, the pain that my criticism would cause the agent criticized if taken to heart, in cases where I believe her to be psychologically normal. As Fischer, Ravizza and Copp write, quoting the Talmud: you cannot chop the head of a chicken and regret its death.<sup>24</sup> There is a conceptual distinction between on the one hand the parting of the head from the body, on the other hand the chicken's death; the latter is a *consequence* of the former rather than identical with it. Still, they are too closely associated for us to intend the one and merely foresee the other.

The distinction between intended and foreseen is sometimes developed in counterfactual terms; I only foresaw a certain consequence, but did not intend it, if I would have preferred that my intention had been realized without it. Does the case of moral criticism pass this test? I doubt that it does, and this is once again because a full realization of the wrongness of one's conduct is so closely tied to psychological suffering. If I am to imagine

---

<sup>22</sup> Philippa Foot, "The Problem of Abortion and the Doctrine of Double Effect", in *Virtues and Vices and Other Essays in Moral Philosophy* (Berkeley: University of California Press, 1978); John M. Fischer, Mark Ravizza and David Copp, "Quinn on the Double Effect: The Problem of Closeness", *Ethics*, 103 (1993): 707-725; Gerry Johnstone, *Restorative Justice: Ideas, Values, Debates* (Cullompton; Willan Publishing, 2002), pp. 31, 110.

<sup>23</sup> Christine M. Korsgaard, *The Sources of Normativity* (Cambridge: Cambridge University Press), 1996), p. 151.

<sup>24</sup> Fischer, Ravizza and Copp, "Quinn on the Double Effect: The Problem of Closeness".

that the person criticized fully realizes the wrongness of her previous conduct with no accompanying psychological suffering, I have to imagine her psychologically very different from the way she is. I do not believe that critics typically prefer a counterfactual situation where the person criticized is peculiarly emotionless about her past wrongdoing to one where the person criticized realizes that she did wrong in the usual, painful manner, nor that critics ought to have this preference. A person who noted the moral reasons against her previous behaviour and resolved to behave differently in the future with no negative feelings – no remorse, no guilt, not even *sadness* – would be emotionally stunted, and this is hardly preferable to her having a normal painful realization of the wrongness of her conduct. Therefore, it is implausible to claim that we can morally criticize (psychologically normal) agents without intending their psychological suffering. Answerability, just as accountability, justifies the intentional infliction of suffering on other people, and there is no argument here for accountability being more freedom demanding.

#### **4. The argument from harsh punishments**

I have argued that answerability as well as accountability justifies making people suffer by making them realize that they did wrong. I have, however, focused solely on the psychological suffering that an agent experiences as a result of realizing that she has done wrong. Philosophers who argue that we cannot deserve to suffer due to not being accountable for anything we do often seem to have more serious suffering in mind. When Parfit argues that we cannot deserve to suffer, he uses eternal torment in Hell as an example. Galen Strawson likewise writes that we cannot deserve an eternity in Hell or various institutionalized punishments. Erin Kelly does not mention Hell, but harsh prison sentences.<sup>25</sup> The suggestion is that accountability is more freedom demanding than answerability, because

---

<sup>25</sup> Parfit, *On What Matters: Volume One*, p. 271; Strawson, “The Bounds of Freedom”, pp. 452, 457; Kelly, “Doing Without Desert”, p. 183.

accountability justifies the infliction of quite *serious* suffering, beyond the mental pain caused by the realization that one did wrong. Presumably, we can only be justified in inflicting such *serious* suffering on others, if what they did was, in a very deep and ultimate sense, up to them.

However, it is important to note that a wrongdoer's accountability is not *sufficient* to justify a harsh punishment. Punishing her harshly may still be wrong for a whole host of reasons. For instance, Robert Kane argues that we have a standing pro tanto obligation to let everyone live their lives as they please. In some situations, it is impossible to let everyone do what they please; if we let certain dangerous people run loose, their victims will have *their* freedom compromised. In such situations it is morally permitted to restrain or coerce wrongdoers in order to protect the freedom of others, but we must not use any more restraint or coercion than is necessary.<sup>26</sup> Kane is a libertarian, and believes that people can be accountable for what they do, but this belief is perfectly consistent with a rejection of harsh punishments. Accountability sceptic Derk Pereboom presents a list of arguments against a retributivist criminal justice system with harsh penalties for which the existence or non-existence of accountability is actually irrelevant. Firstly, he finds it implausible that the state would have an obligation or even a right to spend its resources on a system designed to give people their just deserts. The state's obligations are plausibly limited to protecting its citizens from serious harm and providing a framework for smooth human interaction.<sup>27</sup> These obligations could justify the state instigating a crime reducing criminal justice system, but hardly a retributivist one. Secondly, a retributivist system with harsh punishments is expensive. Thirdly, some innocents will always end up convicted and suffering. Finally, aside

---

<sup>26</sup> Robert Kane, *Through the Moral Maze* (New York: North Castle Books, 1996), pp. 19-30.

<sup>27</sup> There is obviously much room for interpretation here. A political libertarian and a defender of a fairly extensive welfare state might both agree with Pereboom about the obligations of the state to its citizens, but disagree about how to understand 'serious harm' and 'smooth interaction'.

from these more pragmatic arguments against a harsh retributivist system of state punishment, Pereboom argues that it is plausibly wrong in itself to vent one's vengeful desires by punishing wrongdoers.<sup>28</sup> Throughout his writings, Pereboom consistently presents the ideal ethical life as one of compassion, acceptance and forgiveness rather than anger and vengeance. This last argument and his picture of the ideal ethical life are not just relevant for state punishment, but also for private acts of vengeance and the question of whether a morally perfect God would punish wrongdoers eternally in Hell.

This list of accountability independent arguments against harsh punishments is obviously not meant to be exhaustive. There are a number of ways one might argue for the moral impermissibility of private vengeance, a harsh retributivist system of state punishment or for the thesis that a good God would not subject people to eternal torment. One might therefore counter *the argument from harsh punishments* by simply rejecting the premise that accountability justifies the harsh punishment of wrongdoers. If both accountability and answerability justifies the intentional infliction of a certain kind of psychological suffering, as argued in the previous section, but neither justifies inflicting harsh punishments on people, there is no argument here for accountability being more freedom demanding.

However, a defender of the thesis that accountability is more freedom demanding than answerability might respond to his opponent's denial of the premise in the following way: Yes, the fact that a wrongdoer was accountable for what she did does not, in itself, provide sufficient justification for a harsh punishment. Perhaps harsh punishments are, all things considered, wrong, regardless of whether people can be accountable for what they do. But if a wrongdoer was accountable for what she did, it follows that she had *sufficient moral responsibility* to deserve a harsh punishment, and this fact still provides an argument for

---

<sup>28</sup> Pereboom, "Free Will Skepticism and Criminal Punishment", pp. 62-64; *Free Will, Agency and Meaning in Life*, pp. 158-160.

accountability being more freedom demanding than answerability. If people can be accountable for what they do, the harsh punishment of wrongdoers *would* sometimes be justified *if* it were not wrong for other reasons. Now, what kind of free will does accountability, in light of this fact, require? Galen Strawson suggests an argument along these lines. He writes that if people are accountable for what they do,<sup>29</sup> it is “perfectly intelligible” although “morally repugnant” to propose that wrongdoers should be tormented in Hell for all eternity.<sup>30</sup> The supposed fact that accountability would make Hell “intelligible” is meant to support the thesis that accountability requires an utterly impossible kind of free will.

However, I am uncertain whether this argument even makes sense. If subjecting accountable wrongdoers to eternal torment in Hell is morally repugnant – in what way is it still *perfectly intelligible*? Strawson clearly believes that it would be *unjust* to subject someone to eternal torment if she were not accountable for what she did. But if it would be, as he claims, *morally repugnant* to subject her to eternal torment even if she *were* accountable, would it not still be unjust to do so? (Does it even make sense to say that something is simultaneously just and morally repugnant?) If it would still be unjust to do so, how do we get an argument for accountability being particularly freedom demanding from these musings on Hell and eternal torment?

Let us suppose, however, for the sake of argument, that it does make sense for someone who believes Hell to be morally repugnant to ask what kind of free will accountability would require in light of the fact that accountable wrongdoers *would* deserve to go to Hell *if* Hell had been a good idea, and that it makes sense for someone who is opposed to harsh retributivist Earthly punishments to ask what kind of free will accountability would require in light of the fact that accountable wrongdoers *would* deserve such punishments *if* they had not been wrong

---

<sup>29</sup> Or, in Strawson’s exact words, “ultimately and truly without qualification responsible and truly and without qualification deserving of praise and blame”.

<sup>30</sup> Strawson, “The Bounds of Freedom”, p. 451.

for other reasons. Even if such questions make sense, I do not believe that we can answer them by consulting our intuitions. If a person already believes that harsh punishments are morally wrong for other reasons, I doubt that these beliefs can be safely put aside while considering what kind of free will harsh punishments require. It is one thing to ask what would be the right thing to do in a hypothetical situation where the empirical facts are radically different from reality, and quite another to ask what would be right or justifiable if we assume a different ethics. It is one thing to ask oneself “if we imagine that some people have telepathic powers, could there be situations where it would be right for them to read another’s mind without their consent?” It is quite another thing to ask “if we imagine that bullying is not wrong in itself, ought school children to bully those who are different for the fun of it, or ought they to abstain from doing so in order to better focus on their studies?” I seriously doubt that it is possible to generate a reliable intuition about the latter case. I want to answer that they ought to abstain from bullying in order to better focus on their studies, but this is hardly because I have intuited that studies outweigh fun in this situation. Rather, since I believe that bullying is wrong, regardless of what is stipulated in the question, I believe that the children obviously ought not to bully, and then rationalize this conclusion. Likewise, a philosopher who believes that harsh punishments are wrong for a number of reasons, and then asks himself “what kind of free will must wrongdoers have for the justification of harsh punishments, if we imagine that harsh punishments are not morally wrong for independent reasons?” might be lead to answer “an impossible kind of free will; therefore, harsh punishments can never be justified” because he believes such punishments to be wrong regardless of what is stipulated in the question, and rationalizes this conclusion. If we cannot trust our intuitions on this matter, we should suspend our judgement.

But perhaps we should trust, instead, the intuitions of philosophers who *do* believe that there are no conclusive moral responsibility and free will independent reasons to oppose harsh

punishments? Such philosophers, however, sometimes have the intuition that an agent can be accountable for what she does, and thus deserve harsh punishments, as soon as she has a fairly simple compatibilist kind of free will.<sup>31</sup> Likewise, people who do *not* find the idea of eternal torment in Hell for sinners to be morally repugnant for independent reasons have argued that people can deserve this despite lacking libertarian free will. Famous churchmen like Augustine, Martin Luther and Jean Calvin believed that we are born sinful, will remain sinful unless God decides to extend his grace to us and make us better, are utterly incapable of bootstrapping ourselves out of sin, and yet deserve an eternity in Hell if we do not morally improve.

Ultimately, I do not believe that the intuitions of people who embrace harsh punishments are any more reliable than the intuitions of people who oppose them. Someone who wants to embrace harsh punishments for independent reasons might be as motivated to conclude that their justification merely requires a simple kind of compatibilist free will that obviously exists, as the opponent of harsh punishments is motivated to conclude that their justification requires an unlikely or impossible kind of free will.

Whereas I have shown that *the argument from the intentional infliction of suffering* fails to support the thesis that accountability is more freedom demanding than answerability, I have merely thrown doubt on *the argument from harsh punishments*. There are, at least, fairly good reasons to believe that it is wrong to inflict harsh punishments on wrongdoers even if they are accountable for what they do, that a more hypothetical version of the argument does not make sense, and that we cannot, in any case, generate reliable intuitions on the matter. Both

---

<sup>31</sup> Michael Moore, *Placing Blame: A Theory of the Criminal Law* (Oxford: Oxford University Press, 1997), pp. 60-62.

answerability and accountability justifies the intentional infliction of a certain kind of psychological suffering, and accountability does not obviously justify anything more than that.

However, it might still be argued that although the infliction of suffering can be justified by the benefits of moral improvement, protection of victims and improved relationships, it cannot possibly be justified *because the agent deserves to suffer* unless the agent has libertarian free will. Therefore, the more consequentialist answerability still requires less freedom than the non-consequentialist accountability. I will now turn to that argument.

### **5. The argument from a lack of consequentialist justification**

Some philosophers argue that accountability is more freedom demanding than answerability simply because it is a non-consequentialist, backward-looking kind of responsibility.<sup>32</sup> Now, in order for this discussion to make sense, we must firstly assume that the true or best ethical theory is not purely consequentialist. If no action can be morally justified unless it has beneficial consequences, this must be true of blaming people as well. Thus, if pure consequentialism is true, no one can *deserve* blame in a backward-looking sense, regardless of what kind of free will people have. In this paper, I assume that it is possible in principle for actions to have non-consequentialist justifications. With that assumption in place, is a backward-looking kind of moral responsibility necessarily more freedom demanding than a forward-looking one?

Before moving on, however, I should note that it has been argued that blame is justified by the prospects of moral improvement, just as moral criticism is. Antony Duff, who conflates accountability and answerability, argues that the goal and justification of both blame and retributivist punishment is the moral improvement of the blamee or offender, but since this goal is conceptually rather than contingently related to the means, we should not consider the

---

<sup>32</sup> Pereboom, *Free Will Skepticism and Criminal Punishment; Free Will, Agency and Meaning in Life*, Ch. 6; Parfit, *On What Matters: Volume One*, Ch. 11.

justification a consequentialist one.<sup>33</sup> If Duff is right, the argument that accountability is more freedom demanding because it is backward-looking falls. I agree that there seems to be a conceptual connection between sincerely blaming someone and intending, at least insofar as this is possible at all, that she changes her future behaviour in light of the realization that she did wrong. If I sincerely blame someone and thereby cause her psychological suffering, it will disturb me to find out that she typically forgets about blame soon after and continues doing the kind of thing for which I blamed her. If my blame was sincere, I will not shrug my shoulders and say to myself that since I intended to give her the psychological suffering that she deserves, and she did suffer from my blame, all is well. However, accountable agents can obviously deserve to be blamed even in cases where a behavioural change is highly unlikely or even impossible (e.g., someone might deserve to be blamed for an earlier action even if she is on her deathbed, and therefore will not have the opportunity to repeat the behaviour again). If we try to justify these instances of blame by appealing to the value of upholding the general practice of blaming, we seem to have landed squarely in consequentialist territory. Thus, I think we ought to accept that whereas moral criticism is justified by beneficial consequences, blame in the accountability-sense does not need beneficial consequences for its justification.

It might be tempting to say that what justifies us blaming someone in the accountability sense is the fact that she *deserves* to be blamed. However, ‘desert’ is typically seen as a three-part *relation* between a subject and an object that the subject ought to (at least pro tanto) receive in virtue of some fact about the subject.<sup>34</sup> In the case of deserved blame for actions, there is an agent who ought to (at least pro tanto) receive blame in virtue of some fact about

---

<sup>33</sup> Duff, *Trials and Punishment*, p. 7.

<sup>34</sup> Louis P. Pojman, “Does Equality Trump Desert?”, in Owen McLeod and Louis Pojman (Eds.) *What Do We Deserve? A Reader on Justice and Desert* (New York: Oxford University Press, 1998), pp. 283-297; Geoffrey Cupit, “Desert and Responsibility”, *Canadian Journal of Philosophy* 26 (1996): 83-99; Owen McLeod, “Desert”, <http://plato.stanford.edu/entries/desert/> (First published Tue May 14, 2002; substantive revision Wed Nov 12, 2008).

her. Saying that she deserves to be blamed is thus equivalent to saying that there is some fact about her that (pro tanto) justifies us blaming her. It is this *fact* that serves as justification, whereas ‘desert’ is just the name of the relation between the fact, the agent and the blame. What fact, then? We might want to say “the fact that she knowingly performed a wrongful action”, and perhaps add something about it being fitting or apt to respond in certain ways to certain actions.<sup>35</sup> Alternatively, we might say that blaming her is the *respectful* thing to do. Blaming her means treating her like a fellow agent (rather than a child or an animal to be manipulated or trained).<sup>36</sup> It has also been argued that blame is a way of standing up for oneself; that it is connected more to self-respect than respect for blamees.<sup>37</sup> All these facts, if indeed they are facts, might be considered the fact in virtue of which the agent ought to receive the blame. The question is whether we have reason to believe that any of these non-consequentialist justifications require that the blamee had libertarian free will. If blame is justified because it expresses respect, it is hard to see why that would be the case. If we accept that intentionally causing a person pain can be morally right if valuable consequences follow,<sup>38</sup> even if she lacks libertarian free will, why not also accept that intentionally causing her pain can be morally right if doing so realizes some *non-consequentialist* value?

However, if we suppose that blame is justified merely because it is a *fitting* response to certain actions, there is not much to actually *argue* about; we can only consult our intuitions in order to determine what kind of free will an agent must have in order for blame to be a

---

<sup>35</sup> Matt King, “Moral Responsibility and Merit”, *Journal of Ethics and Social Philosophy*, 6 (2012): 1-17; Fischer and Ravizza, *Responsibility and Control. A Theory of Moral Responsibility*, pp. 7-8.

<sup>36</sup> Duff, *Trials and Punishment*, pp. 6, 70; Dworkin, *Justice for Hedgehogs*, pp. 210-211, 224-225; Moore, *Placing Blame*, pp. 142-149, 165.

<sup>37</sup> Matthew Talbert, “The Significance of Psychopathic Wrongdoing”, in Thomas Schramme (Ed.) *Being Amoral: Psychopathy and Moral Incapacity* (Cambridge MA: MIT Press, 2014), pp. 275-300.

<sup>38</sup> Pereboom, *Free Will, Agency and Meaning in Life*, pp. 134-135.

fitting response to her wrongdoing. In the next section, I will discuss thought experiments designed to elicit incompatibilist intuitions on this matter.

## **6. The argument from manipulation cases**

A number of philosophers argue that there are thought experiments that show that accountability is incompatible with determinism. These philosophers may be libertarians or accountability sceptics. Insofar as they are sceptics, they typically believe that answerability, as I have defined it, or something very much like it, is not threatened by said thought experiments. Pereboom, however, has *explicitly* stated that his famous *four-stage manipulation case* is supposed to show that accountability, but not answerability, is incompatible with determinism.<sup>39</sup> I will argue that insofar as the four-stage manipulation case really does threaten accountability, it might very well drag answerability with it. I assume that if this is the case for a thought experiment explicitly designed to threaten accountability only, it will be true of other thought experiments more loosely targeted at “moral responsibility” as well.

In the four-stage manipulation case, there is a gradual move from direct manipulation to more indirect but still determining causes. Pereboom argues that every difference between the scenarios is morally irrelevant. Therefore, the agent lacks accountability in the ordinary deterministic setting if he does so in the first. The main character of the four-stage manipulation case is Professor Plum, who murders Ms White for selfish reasons. Plum can deliberate rationally about what to do, respond to reasons, is selfish but not compulsively so, and is overall described so that he will satisfy the most robust compatibilist freedom requirements. In the first scenario, Plum has a device implanted in his brain, through which

---

<sup>39</sup> Pereboom, *Free Will, Agency and Meaning in Life*, pp. 74-82, 136.

evil scientists directly manipulate him. In Pereboom's first version of the first scenario, Plum was manipulated to think and act as he does moment to moment by the scientists.<sup>40</sup> The problem with this version is that Plum seems to be a mere puppet, and his supposed qualities of being rational, reason-responsive etc. seem like qualities belonging to the scientists rather than to him. Pereboom therefore developed a new version in which the scientists merely push a button that causes Plum to reason in an egoistic way (this is not out of character for Plum, who often reasons in an egoistic way).<sup>41</sup> However, I find it very hard to see how the button-push could be accountability-undermining, given that we, in real life, often are caused to be temporarily more egoistic by various events – we might be temporarily more egoistic if our favourite sports team just lost a game or we stubbed our toes on the furniture – and this is not typically seen as accountability-undermining.<sup>42</sup> If we stress that Plum was *determined* to reason in an egoistic manner by the button, and add that determinism rules out accountability, we would be assuming what ought to be proved. I will therefore move on to the second scenario instead, which is also supposed to be a fairly obvious case of no accountability. In this second scenario the scientists have programmed Plum at an earlier time to reason and deliberate the way he does, and all his deliberations eventually lead him (as the program dictates) to become a murderer.<sup>43</sup> In this scenario, I do feel the pull of incompatibilist intuitions.

There are dyed-in-the-wool compatibilists who do not; philosophers who, when presented with this scenario, merely shrug their shoulders and say that we are all programmed in a sense, by our genes and environmental influences, and why would that be problematic?<sup>44</sup>

---

<sup>40</sup> Pereboom, *Living Without Free Will*, pp. 112-113.

<sup>41</sup> Pereboom, *Living Without Free will*, p. 76.

<sup>42</sup> Pereboom even mentions this in the text.

<sup>43</sup> Pereboom, *Living Without Free Will*, pp. 113-114; *Free Will, Agency and Meaning in Life*, pp. 77-78.

<sup>44</sup> John M. Fischer in conversation.

However, although I ultimately doubt that these incompatibilist intuitions are reliable, I do feel the pull of them when reflecting on this case. They are pumped by the idea that Plum has been *programmed*. If I focus on this description of the situation, rather than on the fact that he is stipulated to satisfy all responsibility conditions that a compatibilist might ask for, I do get the feeling that he is akin to a biological robot. A very sophisticated one, of course, but still a sort of robot that merely obeys its program – and it can obviously not be accountable for *that*. However, as long as I regard Plum this way, I also tend to regard him as lacking *answerability* for what he does.

Plum's behaviour is, of course, not supposed to be rigid. It might be possible to influence his future behaviour by moral criticism. But is it possible to *morally improve* him? Does an agent automatically count as *morally improved* because he changed his behaviour in response to moral criticism? Let us imagine that the IT department at my university played a prank at us moral philosophers, by installing a very particular bug in our computers. From now on, our computers will delete files and in other ways hinder or work unless we write down certain arguments on the keyboard, explaining to the computers the moral wrongness of hindering us. The computers “respond to reasons” in a superficial sense, but we do not find them answerable for their “behaviour”, since they are, after all, mere machines. When I focus on the fact that Plum is *programmed* to behave the way he does, that all his actions are ultimately the result of events over which he lacks control, links in long causal chains running through him back to the scientists, I get the feeling that Plum is analogous to my computer in this thought experiment. Firstly, he performed harmful behaviours. With the right kind of input, the right kind of words, his thoughts might change. He might think thoughts about moral wrongness, and hopefully begins to behave in a more beneficial manner. But how can I legitimately call him *morally improved* when I know that all that happened is that new input resulted in new output? He may be more pleasant to be around once his behaviour has

changed in a more beneficial direction, but how can I say that our *relationship* has improved? How much of an honest relationship can I have with someone of whom I know that he merely acts the way he does because of a combination of early programming and new input?

An agent is answerable when morally criticizing her makes sense because it may contribute to her moral improvement, restoring and improving relationships and protecting other people from future wrongdoings. However, as long as I regard Plum as a kind of biological robot who merely responds to new input with new output, it seems disingenuous to talk of him as morally improved. More pleasant to be around, perhaps, less harmful, but it seems strange to make a *moral* evaluation of an entity that merely combines its initial programming with new input to generate behaviour. Likewise, it is hard to see how I could have an honest relationship with him while regarding him in this light. Protecting other people from future wrongdoing ends up being my only justification for morally criticizing him, and this alone arguably does not suffice for answerability.

But, someone objects, Plum is *not* a biological robot! He is a rational agent. If he changes his behaviour in light of the realization that he did wrong, this is something more than his original programming combined with new input generating new behaviour. You forget all the qualities he has that my computer, in the IT department prank example, lacks! True enough. So let us focus, instead, on all Plum's agential qualities. Let us focus on his rationality, on the fact that he *understands* reasons, and on all the ways in which he is just like you and me. He certainly seems answerable now! And, I will add, he also seems accountable now. If we focus on the fact that he is a rational agent, just like you and me, he certainly seems like a prime candidate for blame after having murdered Ms White for selfish reasons.

Other philosophers have already noted that intuitions about the moral responsibility of agents in thought experiments might oscillate back and forth between compatibilism and

incompatibilism depending on what we focus on.<sup>45</sup> Now, intuitions obviously differ between people as well. As I have already pointed out, some compatibilists never feel the incompatibilist pull, no matter how they focus, and some incompatibilists might likewise be unable to generate compatibilist intuitions. For this reason, there are no knock-down arguments from thought experiments. I have found that for my own part, my answerability intuitions follow my accountability ones. In order to generate accountability incompatibilist intuitions, I must focus on certain features of the scenarios to the expense of others. When I do so, I also feel the pull towards denying Plum's answerability for the murder, even though I still see purely instrumental reasons for providing him with moral arguments akin to the reasons I had in the IT department prank example to "argue" with my computer. When I remind myself that he is a real, rational agent, my answerability intuitions bounce back, and bring accountability with them. Although not providing a knock-down argument, I have shown that it is far from obvious that the four-stage manipulation case targets accountability exclusively.

## 7. Conclusion

The idea that there are different kinds of moral responsibility, some of which are more freedom demanding than others, have seemed plausible to many philosophers. I argue that we can make a meaningful distinction between *answerability* that justifies *moral criticism* for wrongful actions, and *accountability* that justifies *blame*. However, the distinction does not, all things considered, give us reason to believe that accountability is more freedom demanding than answerability. Accountability justifies intentionally making people suffer, but

---

<sup>45</sup> Michael McKenna, "A Hard-line Reply to Pereboom's Four-case Argument", *Philosophy and Phenomenological Research*, 77 (2008): 142-159; "Resisting the Manipulation Argument: A Hard-liner Takes It on the Chin", *Philosophy and Phenomenological Research*, 89 (2014): 467-484; Gunnar Björnsson and Karl Persson, "The Explanatory Component of Moral Responsibility", *Noûs* 46 (2012): 326-354.

so does answerability. Neither accountability nor answerability suffices to justify harsh punishments. Accountability justifies blame for non-consequentialist reasons, but there is no reason to suppose that non-consequentialist reasons require a different kind of freedom than consequentialist ones. There are thought experiments that provoke the intuition that accountability is incompatible with determinism, but the same thought experiments provoke incompatibilist intuitions about answerability. In the absence of successful arguments to the contrary, it is reasonable to assume that accountability requires no more freedom than answerability.