Accountability, Answerability, and Freedom

Sofia Jeppsson

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 nonconsequentialist, and has been 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

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.[1] Gary Watson argues that responsibility has "two faces," one of aretaic evaluation and one of accountability.[2] Robert Kane writes that ultimate moral responsibility requires libertarian free will, and is different from compatibilist deconstructions of responsibility.[3] David Shoemaker argues that our actual moral responsibility practice embodies three distinct conceptions: attributability, answerability, and accountability.[4] Derk Pereboom distinguishes moral answerability from moral responsibility in the basic desert sense.[5]

---

[1]J.J.C. Smart, "Free-Will, Praise and Blame," *Mind* 70 (1961): 291-306.

[2]Gary C. Watson, "Two Faces of Responsibility," *Philosophical Topics* 24 (1996): 227-48.

[3]Robert Kane, *The Significance of Free Will* (New York: Oxford University Press, 1998).

[4]David Shoemaker, "Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility," *Ethics* 121 (2011): 602-32.

[5]Derk Pereboom, "Free Will Skepticism and Criminal Punishment," in Thomas A. Nadelhoffer (ed.), *The Future of Punishment* (New York: Oxford University Press, 2013), pp. 49-78; *Free Will, Agency, and Meaning in Life* (New York: Oxford University Press, 2014), pp. 127-52.

DOI: 10.5840/soctheorpract201642423

There is no consensus among philosophers in the debate about either terminology or where to draw the line between different responsibility types. In part, this can be explained by the fact that philosophers writing about moral responsibility sometimes attempt to answer different questions and approach the issue from different angles. However, a fairly widespread view is 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/or lived up to certain justified moral demands.[6] A focus on this particular distinction is often motivated by moral responsibility skepticism:[7] many skeptics want to argue that although there is an important sense in which no one is morally responsible for what she does, another, also important, kind of responsibility still exists. These philosophers have advanced 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 all those arguments can be resisted.

Following Shoemaker and to some extent Watson,[8] I will consistently use the term *accountability* for the first kind of responsibility, and following Shoemaker and Pereboom,[9] *answerability* for the second kind. *Desert-entailing moral responsibility* or simply *moral responsibility* are more common than *accountability* in the philosophical literature, but *accountability* has the distinct advantage of both being short and acknowledging that there can be other kinds of moral responsibility.[10] I will

---

[6]Derk Pereboom, *Living Without Free Will* (Cambridge: Cambridge University Press, 2001), pp. 139-40; "Free Will Skepticism and Criminal Punishment," p. 51; and *Free Will, Agency, and Meaning in Life*, pp. 127-52; Bruce Waller, "Virtue Unrewarded: Morality Without Moral Responsibility," *Philosophia* 31 (2004): 427-47; and "Sincere Apology Without Moral Responsibility," *Social Theory and Practice* 33 (2007): 441-65; Shoemaker, "Attributability, Answerability, and Accountability"; Derek Parfit, *On What Matters, Vol. 1* (Oxford: Oxford University Press, 2011), pp. 258-72; Gunnar Björnsson and Derk Pereboom, "Traditional and Experimental Approaches to Free Will and Moral Responsibility," in Justin Sytsma and Wesley Buckwalter (eds.), *A Companion to Experimental Philosophy* (New York: Wiley-Blackwell, 2016), chap. 9.

[7]Often but not always: Shoemaker argues that *psychopaths*, rather than everyone in a deterministic universe, might be morally responsible in the latter but not the former sense for their wrongful actions.

[8]Shoemaker, "Attributability, Answerability, and Accountability"; Watson, "Two Faces of Responsibility." Since Watson makes a distinction between accountability and an aretaic evaluation of agents' characters, rather than between accountability and answerability, his "accountability" comes to include some answerability features as well.

[9]Shoemaker, "Attributability, Answerability, and Accountability"; Pereboom, "Free Will Skepticism and Criminal Punishment."

[10]Since there is no universally adopted terminology, it should be noted that other

provide more precise definitions of these terms in the next section. I will then discuss four important arguments for the thesis that accountability is more freedom-demanding than answerability—or for AMFA, as I will abbreviate this thesis henceforth.

2. Accountability and Answerability Defined

An agent is *accountable* for her action when she deserves to be praised or blamed 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. This is not to say that if an agent is accountable for a wrongful act, we are morally obligated to blame her come hell or high water. Few philosophers would be willing to endorse such an uncompromising position. There is no contradiction involved in believing that an agent deserves to be blamed, but since blaming her would have disastrous consequences, we ought not to do so all things considered. Likewise, we can consistently claim that there are several reasons to engage in blaming behavior towards a wrongdoer; one is that she deserves to be blamed, whereas other reasons are forward-looking. The claim that an agent is accountable for what she did does not entail that future consequences are completely irrelevant to the question of how we ought to treat her. It does entail, however, that in the absence of sufficiently strong counterveiling reasons, blaming her is justified if her action was wrong, and praising her is justified if it was exemplary. A number of skeptical philosophers argue that an agent cannot deserve praise or blame unless she has a libertarian, or perhaps an impossible, kind of free will.[11] Now, there are philosophers who call themselves compatibilists because they do not believe that accountability as here defined is important or worth wanting.[12] Whether

philosophers might use these terms differently from how I do. Pereboom, in *Living Without Free Will*, uses "accountability" in the same way as he later came to use, and as I use, "answerability" (pp. 127-57).

[11]Pereboom, *Living Without Free Will*; "Free Will Skepticism and Criminal Punishment"; and *Free Will, Agency, and Meaning in Life*; Waller, "Virtue Unrewarded" and "Sincere Apology"; Parfit, *On What Matters*, chap. 11; Galen Strawson, "The Bounds of Freedom," in Robert Kane (ed.), *The Oxford Handbook of Free Will*, 1st ed. (New York: Oxford University Press, 2002), pp. 441-60.

[12]Pereboom, *Free Will, Agency, and Meaning in Life*, pp. 2-3.

this is so is an interesting philosophical discussion in its own right, but it falls outside the scope of this paper. I will attack the arguments for AMFA head on, rather than arguing that it is irrelevant whether accountability as here defined requires more freedom than answerability since accountability is a useless concept anyway.

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 behavior, and so on. In short, when an agent is answerable for an action, she is an apt target for moral criticism. The practice of engaging in moral criticism is justified because it serves valuable goals: the criticism might morally improve the wrongdoer as well as the person criticizing her (who might achieve important moral insights through the formulation of her criticism), prevent future wrongdoing, and restore or improve relationships.[13] 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 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 behaviors and an increased frequency of desirable ones. Rather, moral improvement is seen as a coming to understand moral reasons better and changing one's behavior 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 behavior in light of those reasons. Therefore, an agent must be responsive to reasons (and perhaps not just reasons in general, but *moral* reasons too) in order to be answerable for what she does. She must have the general ability to understand moral criticism, and to regulate her behavior accordingly. Pereboom argues that reason responsiveness is not only necessary but sufficient for answerability; as soon as the agent has the ability to respond to reasons in the right way, it makes sense to morally criticize her wrongdoing with the aim of morally improving her, preventing future wrongdoing and improving or restoring

[13]Waller, "Sincere Apology," pp. 456-57; Erin Kelly, "Doing Without Desert," *Pacific Philosophical Quarterly* 83 (2002): 180-205, p. 194; Pereboom, *Living Without Free Will*, pp. 156-57; "Free Will Skepticism and Criminal Punishment," p. 51; and *Free Will, Agency, and Meaning in Life*, pp. 132-34; Parfit, *On What Matters*, p. 261; Björnsson and Pereboom, "Traditional and Experimental Approaches."

relationships. He points out that no similarly obvious connection exists between reason responsiveness and accountability.[14] I agree. Some compatibilists[15] argue that reason responsiveness suffices for accountability as well, 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, but to criticize arguments for the thesis that accountability demands more freedom than answerability.

My counterarguments do depend heavily on answerability being defined in a way that makes it a fairly robust kind of responsibility as well. Arguments for AMFA would be much more difficult to resist if AMFA said that accountability is more freedom-demanding than an extremely watered-down version of responsibility. Suppose that an agent counts as morally improved as soon as her ratio of desirable over undesirable actions increases, and that our relationship with her counts as improved as soon as her ratio of desirable over undesirable actions directed at *us* likewise increases. Even if her grasp of reasons (in general, or at least moral ones) is very weak, moral criticism might morally improve her and our relationship with her in this thin sense if she is susceptible to social pressure, and therefore deterred from performing a certain action again after she has been criticized for it. If we define answerability so that agents are answerable for their actions when moral criticism has the potential to improve them and our relationships in this watered-down sense, it will be easy enough to argue that accountability demands more of agents than does answerability. But this is not how accountability-skeptical defenders of AMFA tend to understand answerability, and for good reasons. Unless accountability skepticism still allows for the existence of a fairly robust kind of responsibility, it might be considered too counterintuitive to be plausible, or at least too unattractive to openly embrace.[16] For an accountability skeptic, there are thus good reasons to distinguish between accountability on the one hand and a fairly robust answerability on the other. Since it is the skeptics' arguments for AMFA that I address, I follow them in this.

So far, I have written about blame *or praise*. However, from now on, I will focus exclusively on wrongful actions, blame, moral criticism, and also punishment, rather than their positive analogues. I do so in order to

---

[14]Pereboom, *Free Will, Agency and Meaning in Life*, p. 136.

[15]See, for instance, John Martin Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* (New York: Cambridge University Press, 1998).

[16]As argued by Peter F. Strawson in "Freedom and Resentment," reprinted in Paul Russell and Oisín Deery (eds.), *The Philosophy of Free Will* (New York: Oxford University Press, 2013), pp. 63-83, at pp. 71-73; see also Saul Smilansky, *Free Will and Illusion* (New York: Oxford University Press, 2000).

sidestep the debate about whether blame is more freedom-demanding than praise.[17] 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 the terminology in Pereboom's *Free Will, Agency, and Meaning in Life*, where he uses the term "blame" for both.

I provided, in the introduction, 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.[18] 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 behavior that incurred blame was wrong. I will, however, grant my opponents that blame is different from moral criticism, and that, since accountability justifies blame whereas answerability justifies moral criticism, accountability is different from answerability as well. This is not to say that it is possible to classify all or even most real-life instances of one agent telling another that she has done wrong as belonging to either the category of blame or that of moral criticism. Philosophers who believe that agents can be answerable but never accountable for what they do, that moral criticism but never blame can be justified and that we therefore ought to give up the latter, are not committed to this claim either. It is sufficient for their purposes that blame can be distinguished from moral criticism in theory, and that it would be possible to cultivate a practice of pure moral criticism devoid of blame if people became convinced of the truth of accountability skepticism.

I will accept for the purposes of discussion that blame differs from moral criticism in two important ways. First, blame is accompanied by

---

[17]See, for instance, Susan Wolf, *Freedom Within Reason* (New York: Oxford University Press, 1990), pp. 79-81, and Dana K. Nelkin, *Making Sense of Freedom and Responsibility* (New York: Oxford University Press, 2011), pp. 98-116.

[18]R.A. Duff, *Trials and Punishment* (Cambridge: Cambridge University Press, 1986), pp. 40-41; Ronald Dworkin, *Justice for Hedgehogs* (Cambridge, Mass.: Harvard University Press, 2011), pp. 210-25; Angela M. Smith; "Control, Responsibility, and Moral Assessment," *Philosophical Studies* 138 (2008): 367-92; and "Attributability, Answerability, and Accountability: In Defense of a Unified Account," *Ethics* 122 (2012): 575-89.

certain reactive attitudes such as resentment, indignation, and moral anger.[19] Moral criticism, as defined by philosophers who distinguish accountability from answerability, need not involve an expression of these attitudes. Second, we have already established that accountability, and thus blame, are backward-looking. If blame can be justified at all, it is possible to justify it without appealing to beneficial consequences. Moral criticism, on the other hand, is justified by the goals of moral improvement, prevention of future wrongdoing, and restoration or improvement of relationships. Since blame differs from moral criticism in these ways, and since accountability justifies blame whereas answerability justifies moral criticism, accountability is different from answerability.

At this point, *accountability*, *answerability*, *blame*, and *moral criticism* have been sufficiently well defined to allow for a discussion of the arguments for AMFA. 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. We try to make her suffer—perhaps only psychologically, but psychological suffering is still suffering.[20] It has been argued that wrongdoers can only deserve to suffer if what they did was, in a very strong sense, up to them. Since accountability justifies the intentional infliction of suffering on wrongdoers, it is more freedom-demanding than answerability.[21]

---

[19]Pereboom, *Living Without Free Will*, p. 208; "Free-Will Skepticism and Meaning in Life," in Robert Kane (ed.), *The Oxford Handbook of Free Will*, 2nd ed. (New York: Oxford University Press, 2011), pp. 407-25, at p. 422; *Free Will, Agency, and Meaning in Life*, p. 128; P.F. Strawson, "Freedom and Resentment"; Susan Wolf, "Blame, Italian Style," in R. Jay Wallace, Rahul Kumar, and Samuel Freeman (eds.), *Reasons and Recognition: Essays on the Philosophy of T.M. Scanlon* (New York: Oxford University Press, 2011), pp. 332-47.

[20]Wolf, "Blame, Italian Style," p. 338.

[21]Explicitly argued in Parfit, *On What Matters*, pp. 264-72, and also Wolf, "Blame, Italian Style," pp. 341-44, although she is not a skeptic. Also suggested in: Pereboom, *Living Without Free Will*, pp. 208-9; Bruce Waller, "Denying Responsibility Without Making Excuses," *American Philosophical Quarterly* 43 (2006): 81-90; T.M. Scanlon,

We may, of course, silently blame people in our heads or blame the dead, in which case we do not try to make the blamees feel bad. Neither do we typically intend this if we blame a third party who is not present in the discussion (although this kind of blame might have indirect adverse effects on the blamee later on). Plausibly, though, these kinds of blame are parasitic on the paradigm case of explicit blame directed at the blamee. Furthermore, even silent blame, blame of the dead, or blame of a third party 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 to her face.

All of the above is controversial. There are philosophers who analyze blame differently.[22] 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 AMFA.

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.[23] 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 behavior in light of this realization. Since this is painful, moral criticism too attempts to make the person

---

Moral Dimensions: Permissibility, Meaning, Blame (Cambridge, Mass.: Harvard University Press, 2008), pp. 183-85; and Watson, "Two Faces of Responsibility," pp. 238-39.

It is possible to distinguish the thesis that accountable wrongdoers deserve to suffer from the thesis that they deserve to be blamed with the intention of making them suffer (this was pointed out to me by an anonymous reviewer). However, we do not normally make this distinction when thinking about accountable wrongdoers and what they deserve, and the philosophers I discuss do not make it either. I will assume, in this paper, that accountable wrongdoers deserve both to be blamed and to suffer from the blame.

[22]See, for instance, Scanlon, *Moral Dimensions*, and Pamela Hieronymi, "The Force and Fairness of Blame," *Ethics* 18 (2004): 115-48.

[23]Pereboom, *Living Without Free Will*, pp. 205-26; and *Free Will, Agency, and Meaning in Life*, pp. 134 and 186; Waller, "Sincere Apology," pp. 448-51.

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, disappointment, and sadness.[24] 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 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.[25]

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.[26] It may even be possible for certain agents with unusual 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 in which I believe her to be psychologically normal. As John Martin Fischer, Mark Ravizza, and David Copp write, quoting the Talmud, "you can't cut off the head of a chicken and then say you're not responsible for its death."[27] There is a conceptual distinc-

---

[24]Pereboom, "Free-Will Skepticism and Meaning in Life," p. 422; *Free Will, Agency, and Meaning in Life*, p. 134.

[25]Philippa Foot, "The Problem of Abortion and the Doctrine of Double Effect," *Oxford Review* 5 (1967): 1-15, repr. in *Virtues and Vices and Other Essays in Moral Philosophy* (Berkeley: University of California Press, 1978). John Martin Fischer, Mark Ravizza, and David Copp, "Quinn on Double Effect: The Problem of 'Closeness'," *Ethics* 103 (1993): 707-25; Gerry Johnstone, *Restorative Justice: Ideas, Values, Debates* (Cullompton, U.K.: Willan Publishing, 2001), p. 110.

[26]Although for the view that psychological suffering partly *constitutes* a full moral realization of the wrongness of one's previous actions, see Christine M. Korsgaard, *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996), p. 151.

[27]Fischer, Ravizza, and Copp, "Quinn on Double Effect," p. 707.

tion between on the one hand the parting of the head from the body, and 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 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 criticizers typically prefer a counterfactual situation in which the person criticized is peculiarly emotionless about her past wrongdoing to one in which the person criticized realizes that she did wrong in the usual, painful manner, nor that criticizers ought to have this preference. A person who noted the moral reasons against her previous behavior 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 justify 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. Accountability skeptics 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.[28] Galen Strawson likewise writes that we cannot deserve an eternity in Hell, nor to be subjected to institutionalized punishments.[29] Erin Kelly does not mention Hell, but the extensive use of prison and

[28]Parfit, *On What Matters*, p. 271.

[29]Galen Strawson, "The Bounds of Freedom," pp. 451-52 and 457-58.

even death as criminal punishment in the United States.[30] The suggestion is that accountability is more freedom-demanding than answerability, because 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. This is *the argument from (harsh) punishments*: Accountability justifies harsh punishments, or at least punishment beyond blame; it therefore demands more freedom than answerability, which merely justifies moral criticism.

The argument, as I have described it so far, is a bit vague. I will clarify it in three different ways, creating three slightly different versions of it—all of which have some serious problems.

4.1. The argument from harsh punishments

The first version of the argument goes like this: If agents can be accountable for their wrongdoing, harsh punishments can be justified. Harsh punishments can only be justified if what the agent did was, in a very strong way, up to her. Therefore, accountability is more freedom-demanding than answerability.

This version of the argument can be countered by a denial of the first premise. One might argue that harsh punishments cannot, for reasons quite independent of what kind of free will agents do or do not have, be justified. When it comes to harsh punishments, it is not merely the case, as with blame, that it might be wrong to inflict this on someone if, for example, doing so would have disastrous consequences in this particular situation. One might quite consistently argue that it is *always* wrong to inflict harsh punishments on people, even though accountability exists, and many wrongdoers are accountable for what they do. For instance, accountability libertarian 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.[31] Accountability skeptic Derk Pereboom presents a list of arguments against a retributivist criminal justice system with harsh penalties for which his skepticism is

[30]Kelly, "Doing Without Desert," p. 183.

[31]Robert Kane, *Through the Moral Maze* (New York: North Castle Books, 1996), pp. 19-30.

actually irrelevant: the arguments might just as well be embraced by someone who believes in accountability. First, he finds it implausible that the state has 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.[32] These obligations could justify the state instigating a crime-reducing criminal justice system, but hardly a retributivist one. Second, a retributivist system with harsh punishments is expensive. Third, some innocents will always end up convicted and suffering. Finally, aside 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.[33] 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, for 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 justify the 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 for AMFA here.

A defender of AMFA might, however, attempt to save the argument by tweaking it in a more hypothetical direction. I will now turn to that version.

4.2. The argument from hypothetical harsh punishments

The defender of AMFA might respond to my arguments in the previous subsection in the following way: Yes, the fact that a wrongdoer was accountable for what she did does not, in itself, provide sufficient justifica-

[32]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."

[33]Pereboom, "Free Will Skepticism and Criminal Punishment," pp. 62-64; and *Free Will, Agency, and Meaning in Life*, pp. 158-60.

tion 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 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,[34] it is "perfectly intelligible" although "morally repugnant" to propose that wrongdoers should be tormented in Hell for all eternity. 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, this more complicated and hypothetical version of *the argument from (harsh) punishments* has two big problems: first, it is doubtful whether it even makes sense, and second, it is also doubtful whether we have any reliable intuitions about this matter even supposing that the argument makes sense in the first place. I will begin by discussing the first problem, before I move on to the second one.

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 for other

[34]Or, in Strawson's exact words, "ultimately and truly without qualification responsible and truly and without qualification deserving of praise and blame" ("The Bounds of Freedom," p. 451).

reasons. Even if such questions make sense, I do not believe that we can answer them by consulting our intuitions, and herein lies the second problem for this version of the argument. 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 in which 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 in which it would be right for them to read another's mind without her 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 led 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 judgment.

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 long as she has a fairly simple compatibilist kind of free will.[35] 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

[35]E.g., Michael S. Moore, *Placing Blame: A Theory of the Criminal Law* (Oxford: Oxford University Press, 1997), esp. chaps. 1 and 9.

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.

However, most people believe that we ought to have *some* system of state punishment, and that moderate and humane punishments can be justified. Thus, we get a third version of the argument, omitting the "harsh" part: If agents can be accountable for their wrongdoing, some kind of punishment beyond blame can be justified. Punishments can only be justified if what the agent did was, in a very strong way, up to her. Therefore, accountability is more freedom-demanding than answerability. But even this version has its problems.

4.3. The argument from humane and moderate punishments

It is, of course, still somewhat controversial whether *any* kind of punishment can be justified. Still, most people agree that sufficiently moderate and humane punishments can be justified, and I will assume for the sake of argument that this is the case, so that I cannot simply repeat my reasoning from section 4.1. I have not provided definitions of "moderate" or "humane," and it is compatible with everything I say here that our current system of state punishment is horrible and in need of radical reforms. For the purposes of this paper, however, these issues can be set aside. I assume that moderate and humane—whatever that means more precisely—punishments can be justified, but we must ask ourselves how much, if any, justificatory work is done by *accountability* here.

According to pure retributivism, accountable wrongdoing suffices to justify punishment.[36] Above, I presented a number of arguments against a *harsh* retributivist criminal justice system. At least some of those arguments become less relevant for a more humane retributivist system, based on the idea that offenders ought to get the punishment that they deserve, but that they, for a variety of reasons, do not deserve that much in the way of punishment. Still, according to this humane pure retributiv-

[36]E.g., Moore, *Placing Blame.*

ism, a wrongdoer who is accountable for what she did might deserve punishment beyond blame, and to suffer more than she would from blame alone; the humane pure retributivist can use this as an argument for AMFA. However, pure retributivism, whether it comes in a harsher or more humane version, is highly controversial. Quite independently of the free will issue, many philosophers doubt that making someone suffer can be good in itself.

Even some philosophers who call themselves retributivists (e.g., Jeffrie Murphy, Antony Duff) believe that punishing people requires more justification than "she was accountable for doing wrong," and all non-retributivists do. Punishment (at least in some form, which might be very different from what we have in today's society) is argued to be justified because it serves to drive home to the offender that she really did wrong,[37] because it evens out the unfair advantage that criminals would otherwise have over law-abiding citizens,[38] because it can be regarded as a kind of self-defense (of the state or the public at large),[39] or because it serves a deterrent function.[40] Discussing all the problems and merits of these theories lies outside the scope of this paper. What is important for my purposes is that none of these proposed justifications seems to depend on whether wrongdoers are accountable for what they do. Pereboom himself writes that moral education, deterrence, and self-defense theories are in principle compatible with accountability skepticism, although he rejects them for independent reasons.[41] Arguably, the fairness theory of Jeffrie Murphy, John Finnis, and Richard Dagger is compatible with accountability skepticism as well. They regard a society as fair in the relevant sense when the burdens of self-restraint (i.e., obeying the law even when one does not want to) as well as the benefits from living under law are distributed evenly among the citizens, and unfair when some reap the benefits without bearing the burdens. At least prima facie,

---

[37]Duff, *Trials and Punishment*; Herbert Morris, "A Paternalistic Theory of Punishment," *American Philosophical Quarterly* 18 (1981): 263-71; Jean Hampton, "The Moral Education Theory of Punishment," *Philosophy & Public Affairs* 13 (1984): 208-38.

[38]Jeffrie G. Murphy, "Three Mistakes About Retributivism," *Analysis* 31, no. 5 (1971): 166-69; John Finnis, "The Restoration of Retribution," *Analysis,* 32 (1972): 131-35; Richard Dagger, "Playing Fair with Punishment," *Ethics* 103 (1993): 473-88.

[39]Ferdinand Schoeman, "On Incapacitating the Dangerous," *American Philosophical Quarterly* 16 (1979): 27-35; Daniel M. Farrell, "Deterrence and the Just Distribution of Harm," *Social Philosophy and Policy* 12 (1995): 220-40; Anthony Ellis, "A Deterrence Theory of Punishment," *Philosophical Quarterly* 53 (2003): 337-51; Erin Kelly, "Criminal Justice Without Retribution," *Journal of Philosophy* 106 (2009): 440-62.

[40]Kelly, "Criminal Justice Without Retribution"; Ted Honderich, *Punishment: The Supposed Justifications* (London: Hutchinsons & Co., 1969).

[41]Pereboom, *Living Without Free Will*, pp. 161-74; "Free Will Skepticism and Criminal Punishment," pp. 64-72; and *Free Will, Agency, and Meaning in Life*, pp. 161-69.

this idea of fairness does not seem to depend on the existence of accountability. Skeptics have also presented their own theories of justification, explicitly designed to be independent of accountability: Pereboom and Gregg Caruso argue that the importance of protecting potential victims and rehabilitating criminals can justify locking up dangerous individuals.[42] Ben Vilhauer argues that a fair system of punishment is the one we would agree upon behind a veil of ignorance, where we know that there is no free will and no one is accountable for what she does, but do not know what actions we will be caused to do. In this situation, we would instigate a crime preventive scheme, but would not want to risk very harsh punishments in case we become criminals.[43] Kelly's self-defense/deterrence theory is explicitly designed to be independent of moral responsibility and desert as well.[44]

However, even a nonretributivist might argue that accountability, although not sufficient for justifying punishment, is still *necessary*;[45] that deterrence, self-defense, moral education, and/or fairness *combined* with accountability justifies punishment. One might now argue that since we can be justified in punishing (at least when the punishment is moderate and humane enough and serves some important function) accountable wrongdoers but not merely answerable ones, accountability is more freedom-demanding than answerability. I do not believe that there is anything wrong with this argument for AMFA—*if* one is willing to accept the premise that accountability is necessary for the justification of punishment. But it is important to point out that most accountability skeptics are not willing to accept it, and therefore cannot use *the argument from humane and moderate punishments.* Pereboom, Caruso, Vilhauer, and Kelly all argue that a suitably moderate and humane system of punishment can be justified even if no one is accountable for what she does. And they have strong reasons to do so; it has been argued that accountability skepticism, regardless of its theoretical virtues, is not a theory that we can live by, in part because it supposedly does not allow any punishment of criminals.[46] For precisely this reason, Pereboom,

---

[42]Pereboom, *Living Without Free Will*, pp. 174-85; "Free Will Skepticism and Criminal Punishment," pp. 72-75; and *Free Will, Agency, and Meaning in Life*, pp. 169-73; Gregg D. Caruso, "Free Will Skepticism and Criminal Behavior: A Public Health-Quarantine Model," *Southwest Philosophy Review* 32 (2016): 25-48.

[43]Ben Vilhauer, "Persons, Punishment, and Free Will Skepticism," *Philosophical Studies* 162 (2013): 143-63.

[44]Kelly, "Criminal Justice Without Retribution."

[45]See, for instance, Richard L. Lippke, "Some Surprising Implications of Negative Retributivism," *Journal of Applied Philosophy* 31 (2014): 49-62.

[46]Saul Smilansky, *Free Will and Illusion*; and "Hard Determinism and Punishment: A Practical *Reductio*," *Law and Philosophy* 30 (2011): 353-67.

Caruso, Vilhauer, and Kelly reject the idea that accountability is necessary for the justification of punishment; protection, rehabilitation, deterrence, self-defense, or hypothetical agreements behind a veil of ignorance can justify the occasional punishment of offenders even if no one is accountable for what she does. They further argue that important principles such as never to punish innocents,[47] and to abstain from draconian punishments even in situations in which there are pragmatic reasons for them, can be based on ethical principles that are independent of accountability and desert.

4.4. Conclusion: The argument from (harsh) punishments ought to be rejected, in all its different versions

I have not conclusively refuted *the argument from (harsh) punishments* the way I did with *the argument from the intentional infliction of suffering*, but I have pointed out some important problems that the argument suffers in its various versions—problems that make it reasonable to reject it. We have many good reasons to believe that harsh punishments cannot be ethically justified regardless of what kind of free will and moral responsibility people have. It is doubtful whether it makes sense to ask what kind of free will these punishments would hypothetically require if they could be justified, and in any case, we have no reliable intuitions with which to answer such a question. *The argument from moderate and humane punishments* relies on the premise that accountability is at least necessary for the justification of punishment. I will not, in this paper, take a stand on whether it is. However, I noted that accountability skeptics are unwilling to embrace the premise of this argument, since doing so creates other problems for their theory. They want to argue that merely answerable agents can, in fact, be justifiably punished, if the punishment system has beneficial enough goals and the punishments are sufficiently moderate and humane.

Let us, then, leave punishments aside, and return to the infliction of suffering through criticism and blame. It has been 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 answerabil-

[47]One might object that any accountability skeptic who endorses some system of punishment must be in favor of punishing innocents, since to the skeptic, everyone is innocent. This is undoubtedly true *in a sense*, but in this context, "innocent" is to be read as "did not perform the action in the first place," or as lacking mens rea. The skeptics' respective theories of punishment might justify some demand for mens rea, even if it would look different from today.

ity still requires less freedom than the nonconsequentialist accountability. I will now turn to that argument.

5. The Argument from a Lack of Consequentialist Justification

It has been argued that accountability is more freedom-demanding than answerability simply because it is a nonconsequentialist, backward-looking kind of responsibility.[48] As already noted, the claim that accountability is backward-looking is not plausibly interpreted as implying that we ought to blame accountable wrongdoers regardless of the consequences. However, if a wrongdoer is accountable for what she did, we do not need beneficial consequences, either from this particular instance of blaming or from the general practice of blame, for justification; the fact that she did wrong and was accountable for doing so provides sufficient justification for us blaming her in the absence of strong counterveiling reasons.

Of course, in order for this discussion to make sense, we must 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, it is irrelevant whether a wrongdoer has libertarian free will or even some kind of miraculous self-creating power: she still cannot deserve to be blamed in a backward-looking sense. In this paper, I assume that it is possible in principle for actions to have nonconsequentialist 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 justification a consequentialist one.[49] If Duff is right, the argument that accountability is more freedom-demanding because it is backward-looking fails. 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 change her future behavior

[48]Perhaps most explicitly argued by Pereboom, "Free Will Skepticism and Criminal Punishment," and *Free Will, Agency, and Meaning in Life*, pp. 127-52.

[49]Duff, *Trials and Punishment*, p. 7.

in light of the realization that she did wrong. I stress "sincerely," since someone might, for instance, engage in blaming behavior in order to show off what she considers her moral superiority. Although she displays resentment and anger, she is secretly delighted that the blamee keeps doing the thing for which she blames him, so that she can show off over and over again. But when 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 deserve to be blamed even in cases where a behavioral 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 behavior 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.

How, then, do we justify blaming an accountable wrongdoer, if not by appealing to beneficial consequences? Saying that the blamee deserved the blame does not give us the full story, if we accept the common analysis of desert 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.[50] In the case of an accountable wrongdoer, minimally, she ought to receive blame in virtue of the fact that she did wrong and was accountable for doing so. Perhaps we want to add something about it being *fitting* or *apt* to respond in certain ways to wrongful actions.[51] Some philosophers have even more to say; they add 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).[52] It

[50]E.g., Owen McLeod, "Desert," in Edward N. Zalta (ed.), *The Stanford Encylopedia of Philosophy* (Winter 2008 Edition), http://plato.stanford.edu/archives/win2008/entries/desert/; Louis 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. 283-97; Geoffrey Cupit, "Desert and Responsibility," *Canadian Journal of Philosophy* 26 (1996): 83-99, pp. 84-85.

[51]Matt King, "Moral Responsibility and Merit," *Journal of Ethics and Social Philosophy* 6 (2012): 1-17; Fischer and Ravizza, *Responsibility and Control*, pp. 7-8.

[52]Duff, *Trials and Punishment*, pp. 6 and 70; Dworkin, *Justice for Hedgehogs*, pp. 210-11 and 224-25; Moore, *Placing Blame*, pp. 142-49 and 165; Wolf, "Blame, Italian

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.[53] The question now is whether we have reason to believe that any of these nonconsequentialist justifications require that the blamee had a libertarian or impossible kind of 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,[54] 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 *nonconsequentialist* value, such as respect? One might, of course, argue that the value of respect does not justify blame, since blaming someone is *not* a way of showing her respect.[55] I will not delve into this issue in this paper, but merely stress that *if* blame is justified because it expresses respect, there is no reason to assume that blame (and accountability) would be extra freedom-demanding merely because the justification is not a consequentialist one.

However, if we cut respect out of the picture and suppose that blame of accountable wrongdoers is justified merely because they did what they did, or 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 fitting, apt, or deserved merely because the agent did what she did. 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 skeptics. Insofar as they are skeptics, 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

---

Style," p. 339.

[53]Matthew Talbert, "The Significance of Psychopathic Wrongdoing," in Thomas Schramme (ed.), *Being Amoral: Psychopathy and Moral Incapacity* (Cambridge Mass.: MIT Press, 2014), pp. 275-300.

[54]Pereboom, *Free Will, Agency, and Meaning in Life*, pp. 134-45.

[55]Ibid., p. 182.

answerability, is incompatible with determinism.[56] 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 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.[57] The problem with this version is that Plum seems to be a mere puppet, and his supposed qualities of being rational, reason-responsive, and so on 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).[58] 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 favorite sports team just lost a game or we stubbed our toes on the furniture—and this is not typically seen as accountability-undermining.[59] 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

---

[56]See ibid., pp. 74-82, for a description of the case, and p. 136 for the claim that it does not target answerability.

[57]Pereboom, *Living Without Free Will*, pp. 112-13.

[58]Pereboom, *Free Will, Agency, and Meaning in Life*, pp. 76-77.

[59]Pereboom actually mentions, when discussing his new version of the first scenario, that we often become momentarily more egoistic because of some external influence over which we lack control, and that this does not undermine *agency*. He considers this a point in favor of the new version, since no one can accuse *this* Plum of being a mere puppet.

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.[60] 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?[61] However, for most of us who *do* feel the pull when reading Pereboom's description of the case, the pull is dependent on the description and what we focus on. When reading Pereboom's description of how Plum has been *programmed* to act the way he does, I get the feeling that he is, in crucial ways, similar to a sophisticated biological robot. With a different description, with a focus on Plum's agency, reasons for action, and all the ways in which he is just like you and me, the incompatibilist intuitions dissolve. This is, of course, nothing new; other philosophers have discussed in depth how our intuitions tend to oscillate depending on how a case is described and which factors we focus on.[62] Since the purpose of this paper is not to argue for compatibilism or incompatibilism, I will not try to solve the problem of how we ought to describe a case in order to produce *reliable* intuitions. What I will argue is that answerability intuitions might be as unstable as accountability ones; I know that mine are. In order to support AMFA (rather than the thesis that answerability and accountability both require libertarian free will), it is crucial that the manipulation case targets accountability *exclusively*, and this is what I question.

Let us pump our incompatibilist intuitions. Let us focus on how Plum, in stage two, has been *programmed* by the scientists to act as he does. True, he thinks about various reasons pro and con various actions before he makes a decision and acts on it, but every stage of his deliberation follows mechanically from his original programming combined with environmental influences. Is Plum, in this stage, really an apt target for moral criticism? Does it make sense to criticize him in order to improve him morally, improve our relationship with him, and prevent future wrongdoing? Preventing future harmful actions on his part is something we might very well try to do by exposing him to various moral arguments, since this is one kind of input that might change the output ac-

[60]Pereboom, *Living Without Free Will*, pp. 113-14; *Free Will, Agency, and Meaning in Life*, p. 77.

[61]John Martin Fischer in conversation.

[62]Michael McKenna, "A Hard-Line Reply to Pereboom's Four-Case Argument," *Philosophy and Phenomenological Research* 77 (2008): 142-59; and "Resisting the Manipulation Argument: A Hard-Liner Takes It on the Chin," *Philosophy and Phenomenological Research* 89 (2014): 467-84; Gunnar Björnsson and Karl Persson, "The Explanatory Component of Moral Responsibility," *Noûs* 46 (2012): 326-54.

cording to his programming. But can we really morally improve him? In what way has he become *morally improved* and a *better person* just because we managed to change his behavior by giving him new input? Suppose that the IT department at the university decided to play a prank on us moral philosophers by installing a bug in our computers that would cause them to delete important files and in other ways hinder our work, unless we "reasoned" with them by writing down certain moral arguments on the keyboard. Until we had managed to convince the IT department to remove the bugs again, we would be motivated to do so—but we would not really consider the computers as morally improved after we had "reasoned" with them. It is not just that computers cannot be accountable for what they do and deserve blame or praise—their "actions" have no moral value at all. They can "act" in a way that is more or less desirable to us, but not right or wrong, morally good or morally bad. If Plum is crucially similar to a robot or computer, I think the same applies to him. He cannot be morally improved, it is doubtful whether we can talk about preventing future *wrongdoing* (as opposed to merely "future harmful behavior"), and he does not seem like the kind of creature that you can have an actual *relationship* with. He does not merely lack accountability for what he does, but answerability as well.

But, someone objects, Plum is *not* a computer or robot! He is a rational agent. If he changes his behavior in light of the realization that he did wrong, this is something more than his original programming combined with new input generating new behavior. 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.

This is obviously no knock-down argument against AMFA, since it is built on my intuitions. As already noted, intuitions differ. However, I am quite confident that my own intuitions are not unique. Peter van Inwagen, for instance, seems to believe that accountability and answerability (actually, accountability and all of practical reason and morality) stand and fall together.[63] To sum up, the problem with manipulation arguments like Pereboom's four-stage case is not only that some dyed-in-the-wool compatibilists do not feel the pull of accountability incompatibilist intuitions

[63]Peter van Inwagen, *An Essay on Free Will* (Oxford: Clarendon Press, 1983). See, for instance, pp. 154 and 207.

when reading them. The problem is also that when accountability incompatibilist intuitions do arise, they might be incompatibilist about answerability as well. If so, the manipulation argument fails to support AMFA.

7. Conclusion

The idea that there are different kinds of moral responsibility, some of which are more freedom-demanding than others, has seemed plausible to many philosophers. I agree that we can make a meaningful distinction between *answerability* that justifies *moral criticism* for wrongful actions, and *accountability* that justifies *blame*. Furthermore, it has been argued that no one can be accountable for what she does, because we do not have the right kind of free will, whereas many people *are* answerable for their actions, because answerability requires less freedom than accountability. I show that none of the arguments for this thesis is really convincing. Accountability justifies intentionally making people suffer, but so does answerability. Harsh punishments are plausibly unjustified regardless of what kind of moral responsibility and free will people have, whereas it is doubtful whether accountability must play any part in the justification of moderate and humane punishments. Accountability is backward-looking, but that does not seem to give us any reason for assuming that it is more freedom-demanding. 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.[64]

Department of Philosophy, Linguistics, and Theory of Science,
University of Gothenburg
sofia.jeppsson@gu.se

[64]This research was funded by the Anna Ahlström and Ellen Terserus Foundation.