

Moral Responsibility, Praise, and Blame

Hannah Tierney and Robert H. Wallace (equal authorship)

Penultimate draft, please cite published version

1. Introduction

Moral responsibility, praise, and blame are interconnected in a number of ways, but precisely how, and the extent to which, these concepts are related will depend on the details of one's philosophical account. On some views, the connection between moral responsibility, praise, and blame is quite tenuous—it's possible to theorize about one of these concepts without reference to the others. For example, sceptics about free will and moral responsibility will often defend views of praiseworthiness and blameworthiness that do not require agents to be morally responsible in order to be the appropriate target of certain forms of praise and blame (Pereboom 2017, 2021). Still others think that we can hold responsible without blaming (Pickard 2011, 2017). And some theorists take blameworthy and praiseworthy actions to differ along more dimensions than mere valence. As Susan Wolf and Dana Nelkin have argued, agents must have the ability to do otherwise if they are to be blameworthy for performing wrong actions, but do not need this ability in order to be praiseworthy for performing right actions (Wolf 1980, 1990; Nelkin 2011; Brink and Nelkin 2013). But on other views, there is a much tighter conceptual relationship between moral responsibility, praise, and blame—theorizing about one requires theorizing about the others. This is particularly true of Strawsonian views, those that build on P. F. Strawson's work in "Freedom and Resentment" (1962) and analyse being morally responsible in terms of the praising and blaming practices we engage in to hold agents responsible. On these approaches, it is impossible to make sense of what it is to be morally responsible without reflecting on our practices of praising, blaming, and holding agents responsible, and vice versa. In this chapter, we will explore Strawsonian accounts of moral responsibility, praise, and blame, focusing on what these views stand to gain by conceiving of these concepts as so closely connected and what they risk in doing so.

Our plan for this chapter is as follows. In section 2, we explore Strawson's (1962) original discussion of the nature of moral responsibility, highlighting how his focus on the reactive attitudes allows him to defend a compatibilist account of the relationship between determinism and moral responsibility, while also capturing a wide range of our praising and blaming practices. In section 3, we consider a prominent objection to Strawson's argument that determinism is irrelevant to moral responsibility and explore two potential replies. In section 4, we present a critique of Strawson's account of our praising and blaming practices and outline a possible defence. We conclude in section 5.

2. Strawson's Approach to Moral Responsibility, Praise, and Blame

The relationship between the thesis of determinism and moral responsibility has been long contested in the philosophical literature. In order to understand the tension between determinism and moral responsibility, it will be helpful to offer brief descriptions of these concepts:¹

Determinism: The facts of the past, in conjunction with the laws of nature, entail every truth about the future. (Fischer et al. 2007)

¹ The definitions offered here, while well-represented in the literature, are not without controversy. The arguments discussed in this chapter do not hang on the particular ways we've presented these concepts, and it should be possible to re-create these arguments using different conceptions of determinism and moral responsibility.

Moral Responsibility: To be morally responsible is to be the fitting and/or deserving target of moral praise and blame.

Many theorists (e.g., Caruso 2021; Chisholm 1982; Kane 1996; Pereboom 2001; van Inwagen 1983) have argued that the truth of determinism is incompatible with moral responsibility: in a world in which determinism is true, praise and blame would be unfitting and/or undeserved. These incompatibilist philosophers argue that it would be unfair to praise or blame an agent for performing an action that they were determined to perform, since these agents do not have the right kind of control over their action to be the fitting and/or deserving target of praise and blame.² Compatibilists—those who think that the truth of determinism is compatible with moral responsibility—reject the claim that agents are not morally responsible in deterministic universes. In “Freedom and Resentment,” Strawson (1962) presents a novel and influential compatibilist account of moral responsibility, which differs significantly from the views of compatibilists that came before him.

First, rather than focus on how our practices of punishment and moral condemnation interact with the truth or falsity of determinism, Strawson focuses on our “non-detached attitudes and reactions of people directly involved in transactions with each other; the attitudes and reactions of offended parties and beneficiaries, and of such things as gratitude, resentment, forgiveness, love, and hurt feelings” (1962: 75). This is because the practices of punishment and moral approbation allow a certain kind of detachment that these *reactive attitudes* do not. According to Strawson, our disposition to have these reactive attitudes constitutes a moral demand for good will from others, and our praising and blaming practices are scaffolded around these attitudes. Indeed, Strawson contends that this framework of moral emotions is part of what it is to be interpersonally engaged with others. He writes: “in the absence of any forms of these attitudes it is doubtful whether we should have anything that we could find intelligible as a system of human relationships” (1962: 80).

Strawson argues that reflection on the reactive attitudes reveals how very much we care about others’ attitudes and intentions towards us. In particular, we care about whether the actions of others reflect attitudes of good will toward us or whether their actions reflect contempt, indifference, and malevolence. Strawson (1962: 63) asks us to think about the relationships we have with others to see the importance we attach to attitudes of good will and the reactive attitudes we express in light of others’ attitudes. Absent special considerations, we typically feel gratitude towards those who show us good will, resentment and indignation towards those who show us or our loved ones ill will, and guilt when we show others ill will. For Strawson, experiencing and expressing these reactive attitudes forms the bedrock of our practices of praising, blaming, and holding others, as well as ourselves, responsible.³ So, it will be important to examine the special circumstances in which we do *not* experience or express

² For the purposes of this paper, we will remain neutral about what kind of control is necessary for moral responsibility. There is a substantive dispute in the literature between those who understand control in virtue of the ability to do otherwise (Kane 1996; van Inwagen 1983; Vihvelin 2013) and those who ground control directly in terms of the actual sequence of events (Fischer and Ravizza 1998; Pereboom 2001; Sartorio 2016). Note that one could be an incompatibilist or compatibilist about one, both, or neither of these forms of control.

³ In fact, some have read Strawson as arguing that being responsible metaphysically depends on the aptness conditions for holding responsible (Wallace 1994). However, others argue that the best way of understanding the relationship between holding and being responsible is epistemic, such that an investigation into holding responsible gives us evidence about the conditions for being responsible (Brink and Nelkin 2013).

these emotions, since this will give us insight into the conditions under which agents are not morally responsible.

Strawson (1962: 72–73) first discusses special circumstances that involve *excuses* and *justifications*. Imagine that someone knocks you over as you're walking down the sidewalk. Our first reaction might be to blame the person, to resent them, and perhaps to express that resentment through a sharp remark. But if we were to discover that the agent didn't push us intentionally, but rather tripped and fell in our way, our resentment would likely dissipate. Though this individual's action appears to express ill will, the agent in question does not in fact have ill will towards us, and we can see that when they offer their *excuse*. Excuses operate by indicating that the agent in question, though they are a fully developed moral agent and generally morally responsible for their behaviour, is not morally responsible or blameworthy for a particular action because they do not show us ill will in performing that action. Agents can also *justify* their behaviour in ways that alleviate our blaming reactive attitudes towards them. If a person knocks into us on the sidewalk but does so in order to push us out of harm's way, perhaps out of the path of a speeding car, we would surely not resent them. And this is because their action, despite appearances, wasn't wrong at all. Thus, excuses and justifications are attempts to show others that an action that looked like it was motivated by less than good will was not what it first appeared to be. They invite us to see the putative injury in a new light (Strawson 1962: 73).

Another category of special considerations involves *exemptions*. Exemptions are offered when the agent in question doesn't have the relevant capacities to be in the kinds of interpersonal relationships that open agents up to being the appropriate target of our reactive attitudes. Young children are perhaps like this, as are people who are severely mentally ill, or in the late stages of dementia and other neurodegenerative diseases. These agents might in fact show us ill will, but it is wrongheaded to expect that they show us good will (of the pertinent sort), since they are unable to participate in the relationships within which such a demand makes sense. In these cases, we should view the agent in question with *objective* attitudes, rather than reacting to them with attitudes constitutive of participation in interpersonal relations. At the extreme, these objective attitudes involve seeing someone as an object "to be managed or handled or cured or trained" (Strawson 1985: 25). To use Strawson's own terminology, excuses and exemptions are appropriate for normal persons. In contrast, exemptions indicate either temporary or permanent abnormality by way of incapacitation from normal interpersonal relations.⁴ Instead of inviting us to see a putative injury in a new way, exemptions ask us to "see the *agent* as other than fully responsible" (Strawson 1962: 73, emphasis added).

We are now in a position to explore Strawson's argument that determinism is irrelevant to moral responsibility. If the truth of determinism were incompatible with agents being morally responsible, then we should expect it to operate as an excuse, justification, or exemption. But the truth of determinism does not operate as an excuse. Excuses apply to particular agents in particular circumstances, but determinism, if true, would apply to all agents in all circumstances. Furthermore, excuses are successful when they show that the agent in question doesn't really show others ill will even though their action seems to indicate that they do. But the truth of determinism wouldn't tell us anything about an agent's quality of will. Imagine that someone knocked into you on the sidewalk and upon being confronted they responded: "Determinism is true." This gives us no information about the attitudes they have towards others, or whether they intended to perform the action in question, or

⁴ Strawson's use of the term "abnormal" has caused some interpretive controversy. See Bennett (1980) for criticism. For concerns about ableism and Strawson's account of the reactive attitudes, see Ciorria (forthcoming).

whether they knew they were performing such an action. Thus, the truth of determinism does not impact the quality of agents' wills and cannot function properly as an excuse.

The truth of determinism also cannot operate as a justification. Justifications function by showing us that an action has a different moral status than it initially appeared to have. But the truth of determinism doesn't alter the moral status of our actions. Again, imagine that someone knocks into you on the sidewalk and, when confronted, states: "Determinism is true." This would do nothing to change the moral quality of the action itself—it doesn't cause the action to produce the most possible happiness, or the maxim of the agent to be in-line with the categorical imperative, or the action to be helpful, benevolent, or otherwise virtuous. Thus, the truth of determinism cannot function successfully as a justification.

And determinism also fails to fit the shape of an exemption. Exemptions are issued on behalf of agents who do not have the capacities required to engage in everyday interpersonal relationships. This is why they are exempt from the reactive attitudes—they cannot be subject to the demands of good will that our interpersonal relationships involve. But the truth of determinism doesn't make it impossible to be in meaningful relationships with others. Imagine that your spouse comes home one day and tells you: "I have some bad news, determinism is true. It's inappropriate for me to continue having the rich range of reactive attitudes towards you that constitutes our relationship. I want a divorce." That would be nonsensical. The truth of determinism doesn't alter what kinds of cognitive and emotional capacities we have, just as it doesn't change the quality of our wills or the moral status of our actions. Thus, it doesn't affect the kinds of relationships we are able to have and cannot function as an exemption.

Additionally, Strawson argues that it would be practically impossible to treat determinism like an exemption (1962: 74). If determinism were to function like an exemption, then we would have to refrain from experiencing and expressing the reactive attitudes towards others and instead adopt the objective stance towards them. But it would be incredibly difficult and perhaps impossible to take the objective attitude towards everyone all the time. This is because taking the objective stance towards others precludes our ability to engage in the rich and substantive interpersonal relationships that characterize our moral lives. And Strawson claims that these relationships are indispensable: "our inter-personal relationships are too thoroughgoing and deeply rooted for us to take seriously the thought that a general theoretical conviction could change us so that there were no longer anything called interpersonal relationships" (1962: 66).

Once Strawson shows that the truth of determinism cannot function as an excuse, justification, or exemption, he is able to conclude that determinism poses no threat to moral responsibility. The truth of determinism doesn't affect any of the things that make our praising and blaming practices morally appropriate, be it the quality of our wills, the moral status of our actions, or our abilities to engage in rich and meaningful relationships, and thus is entirely compatible with moral responsibility. In coming at the incompatibility problem by way of our feelings and attitudes, Strawson's compatibilism is deeply immodest. For Strawson's argument suggests that the basic moral concerns expressed in our interpersonal moral practices—irrespective of the "local and temporary features of our own culture"—are not threatened by the truth of determinism at all (1962: 80).⁵ Indeed, Strawson argues that even if our moral practices of praise and blame were somehow threatened, our "human commitment" to the concerns behind these practices are impossible to relinquish (1962: 26). Beyond

⁵ See Beglin (2018) for further discussion of a concerns-based understanding of Strawson.

this, given that we have a basic concern for the attitudes of others, it would be practically irrational to give them up even if we could.

3. Can Strawsonian accounts of moral responsibility really avoid the threat of determinism?

While Strawson's work on free will and responsibility has been incredibly influential, many philosophers have resisted his arguments in various ways. In this section, we'll explore one way some have objected to Strawson's argument for compatibilism and a few ways of revising the Strawsonian account in the face of such an objection.

3.a. Manipulation and determinism

Recently, compatibilist views of moral responsibility have been the targets of *manipulation arguments*.⁶ Manipulation arguments come in many forms, but one popular variant has the following structure:⁷

- (1) Manipulation Premise: Manipulated agents are not morally responsible.
- (2) No-Difference Premise: Determinism is relevantly similar to manipulation.
- (3) Anti-Compatibilist Conclusion: So, the truth of determinism is incompatible with moral responsibility. (Latham and Tierney 2022: 144)

Arguments of this kind target a wide variety of compatibilist accounts of moral responsibility, and Strawsonian views are no exception. For example, Derk Pereboom (2014) argues that his four-case manipulation argument undermines Strawsonian approaches to moral responsibility, along with several other compatibilist accounts.

Pereboom develops his argument by asking his readers to consider a series of cases in which an agent, Plum, who meets a variety of compatibilist sufficient conditions for moral responsibility,⁸ is subject to increasingly benign forms of manipulation that causally determine him to decide to kill White. Pereboom begins with the following case:

Case 1: A team of neuroscientists has the ability to manipulate Plum's neural states at any time by radio-like technology. In this particular case, they do so by pressing a button just before he begins to reason about his situation, which they know will produce in him a neural state that realizes a strongly egoistic reasoning process, which the neuroscientists know will deterministically result in his decision to kill White. Plum would not have killed White had the neuroscientists not intervened, since his reasoning would then not have been sufficiently egoistic to produce this decision. But at the same time, Plum's effective first-order desire to kill White conforms to his second-order desires. In addition, his process of deliberation from which the decision results is reasons-responsive; in particular, this type of process would have resulted in Plum's refraining from deciding to kill White in certain situations in which his reasons were different. His reasoning is consistent with his character because it is frequently egoistic and sometimes strongly so. Still, it is not in general exclusively egoistic, because he sometimes successfully regulates

⁶ See Kane (1996), Mele (1995, 2006), Pereboom (2001, 2014).

⁷ This discussion of manipulation arguments draws on Latham and Tierney (2021, 2022).

⁸ For compatibilist accounts of the sufficient conditions for moral responsibility, see Hume (1739), Frankfurt (1971), Fischer and Ravizza (1998), and Wallace (1994).

his behavior by moral reasons, especially when the egoistic reasons are relatively weak. Plum is also not constrained to act as he does, for he does not act because of an irresistible desire—the neuroscientists do not induce a desire of this sort. (Pereboom 2014: 76-77)

Pereboom argues that Plum is intuitively not morally responsible for his decision in this case, despite the fact that he meets several different compatibilist sufficient conditions for moral responsibility. Pereboom then contends that we find Plum to be not morally responsible in this case because his decision was determined by factors beyond his control.

Next, Pereboom asks his readers to consider Case 2, which is almost identical to Case 1 except Plum's decision is determined by a team of neuroscientists who program him at the beginning of his life. Like in Case 1, Plum meets a wide range of sufficient compatibilist conditions for moral responsibility, but Pereboom maintains that he is intuitively non-responsible. And, because Case 2 is relevantly similar to Case 1, consistency requires us to judge that Plum is not responsible in Case 2 just as we judge him to be not responsible in Case 1. Then, Pereboom presents Case 3, which is relevantly similar to Cases 1 and 2, except Plum's decision is determined by the training practices of his community. Again, Pereboom argues that consistency requires us to judge Plum to be non-responsible in Case 3, since it is relevantly similar to Cases 1 and 2. Finally, Pereboom presents Case 4. In Case 4, the thesis of determinism is true: "everything that happens in our universe is causally determined by virtue of its past states together with the laws of nature" (2014: 79). In this case, Plum's decision is determined by facts about the past and the laws of nature. Pereboom argues that consistency requires us to judge that Plum is not responsible in this case, just as he is not responsible in Cases 1–3, since there are no relevant differences between these cases. But if we conclude that Plum is not responsible in Case 4 because determinism is true, then we must issue the same judgment for all agents in deterministic universes. Thus, we must reject compatibilism about moral responsibility if we judge that Plum is non-responsible in Cases 1–4. In this way, Pereboom's argument takes the form introduced above:

- (1) Manipulation Premise: Plum in Cases 1–3 is not morally responsible for his decision to kill White because of the way in which he was manipulated.
- (2) No-Difference Premise: Plum in Case 4 is not morally responsible for his decision to kill White because the truth of determinism featured in Case 4 is relevantly similar to the manipulation featured in Cases 1–3.
- (3) Anti-Compatibilist Conclusion: So, compatibilism is false: Normally functioning agents in deterministic universes, because they are relevantly similar to Plum in Case 4, are not morally responsible for their actions. (Latham and Tierney 2022: 148–149)

Notice that Strawsonian accounts are especially vulnerable to Pereboom's four-case argument because it presents an "internal challenge" to the Strawsonian framework (Pereboom 2014: 74–79). The argument targets the very attitudes and emotions that Strawsonians take to ground moral responsibility. Strawson argues that the thesis of determinism is irrelevant to our practices of holding others morally responsible, but Pereboom's Cases 1–3 indicate that our practices are susceptible to a wide range of exemptions that arise from manipulation. This claim looks even more plausible when one considers the fact that Strawson himself suggests that "peculiarly unfortunate" formative circumstances can exempt agents (1962: 66). If the way an agent was raised can make it appropriate to take the objective stance towards them and exempt them from our praising and blaming practices, then surely the kinds of manipulation featured in Cases 1–3 can have the same impact. But the manipulation in Cases 1–3 is relevantly similar to the thesis of determinism, as it is presented in Case 4. And, Pereboom argues, our practices require us to be consistent: "It is also an internal feature of

the practice that if no relevant moral difference can be found between agents in two situations, then if one agent is legitimately exempted from moral responsibility, so is the other” (2014: 82). Because Case 4 is relevantly similar to Cases 1–3, we should judge that Plum is exempt from being held responsible in Case 4, just as he is exempt in Cases 1–3. If this is right, then determinism is not irrelevant to our practices of praising, blaming, and holding responsible—in fact, it is incompatible with these practices, and given Strawson’s view, moral responsibility itself.

There are two main paths a compatibilist can take when responding to manipulation arguments like Pereboom’s four-case argument. One can take what Michael McKenna (2008) calls the hard-line and deny the Manipulation Premise or one can take what he calls the soft-line and deny the No-Difference Premise. We explore how a Strawsonian might pursue each of these strategies below.

3.b. Response 1: Manipulation does not rule out responsibility

First, consider the hard-line reply, which denies the Manipulation Premise. McKenna (2008) suggests that we can run Pereboom’s cases in reverse while thinking about what a fair-minded, neutral inquirer would say about each. McKenna argues that such a neutral inquirer would be unsure about what to say in Case 4. By parity of reason, they should be unsure about what to say regarding Case 3, Case 2 and Case 1. So, McKenna thinks that there is a case to be made against the Manipulation Premise. In support of this idea, he cites Nomy Arpaly’s use of real-life cases (2003: 127–129). McKenna asks his readers to consider a case in which a woman loses a parent to leukemia at a young age:

Whether for good, rational reasons or not, suppose those experiences settled for that child what would become her deepest unsheddable values about how to live. And suppose that as a mature adult she acts upon them. Does she do so unfreely?... According to her, she regards this not as an impediment of her freedom and her responsibility or, one might say, her dignity, but as a condition of it. Thus was she so made. But as she sees it, it surely does not undermine her free and responsible agency. It makes it. (2008: 156)

On McKenna’s view, even though the woman’s values are completely determined by events entirely outside of her control, they are still very much *her* values, and she can freely act upon them. The same can be said of the radical convert or the person dramatically changed post-accident. Indeed, the closer we get to real-life cases, the more plausible it seems to hold manipulated agents responsible.

But even if this response can put pressure on the Manipulation Premise, it will require the Strawsonian to revise their view and to reject Strawson’s original claim that unfortunate formative circumstances matter for moral responsibility. Unlike Strawson himself, the revisionary view will try to move from an account of our practices of praise and blame to an account of the underlying capacities and abilities needed to be morally responsible, and in doing so, show that these capacities are compatible with the truth of determinism.⁹ This metaphysical explanation would then allow the Strawsonian to hold the hard-line view in a principled manner. We will explore a minimal revision by sticking closely to Strawson’s concern for quality of will.

Consider the fact that Strawson’s account suggests that blameworthy wrongdoing is wrongdoing that manifests ill will, or a lack of sufficient good will, towards others. Indeed, one can argue that what

⁹ Wallace (1994) is the *locus classicus* for this kind of revisionary Strawsonian strategy.

makes a wrong action *blameworthy* is that it displays the absence of sufficient good will.¹⁰ Our reactive attitudes and moral responsibility practices reflect the fact that we care about good quality of will. Drawing on these attitudes and practices, we might say that morally responsible agents—those agents whose actions are apt candidates for praise and blame—are those that can be held to a standard of displaying sufficient good will (or at least no ill will) towards others.

The sort of abilities and capacities needed to be held to a standard of displaying sufficient good will seem to come in three varieties: emotional, volitional, and intellectual. Responsible agents have the emotional capacities needed to have good will towards others. They also have emotional capacities and abilities required to react to others' displays of quality of will with resentment, indignation, gratitude, and other reactive emotions. Moreover, responsible agents are able to act voluntarily in a way that enables their actions to be genuine expressions of good will towards others. While accidental actions can benefit others, they cannot manifest good will. For example, if someone were to accidentally trip, pushing you out of the way of oncoming traffic, this would not be an indicator of their quality of will, even though it clearly benefits you. Thus, the ability to act voluntarily is importantly related to the ability to express quality of will. Finally, responsible agency has an intellectual component. Morally responsible agents understand how to assess the actions of others and can later correct any misunderstandings if offered an excuse or justification. And they also understand that their actions are likewise assessable by others (cf. Russell 2017; McKenna 2012). Thus, it seems like the Strawsonian should adopt a minimal metaphysical commitment: the appropriateness of praise and blame is tied to the suite of interpersonal abilities needed to have and display quality of will. Per our discussion in section 2, the truth of determinism does not appear to entail either that these abilities are always used well or that no agent ever has these abilities.

Notice that on this way of developing the revised Strawsonian position, an agent's causal history is not relevant to whether they are morally responsible for any particular action so long as they have the pertinent abilities to express quality of will. Although an agent's history can explain why they have or lack these abilities, following Strawson's read on our moral practices, praise and blame are made apt by the quality of will displayed in action, and so the causal history of an action only matters insofar as it leads to or does not lead to the presence or absence of manifest quality of will. Given this, we might think it does not matter where in the history of a person's formation good will (or lack thereof) came from. Like Harry Frankfurt (1988: 54), the revisionary Strawsonian could deny that the sources of our abilities are relevant to the control, and so moral responsibility, that those abilities currently afford us.¹¹ Saying something like this would be a natural development of Strawson's argument that puts pressure on the manipulation premise precisely where the hard-liner needs it. Plum is designed in Pereboom's cases to meet compatibilist sufficient conditions for moral responsibility. In this case, that means having all the pertinent abilities needed to meet a standard of displaying sufficient good will. Given this, Plum's being physically determined, or being indoctrinated by his community, or being tampered with by neuroscientists, does nothing to change the fact that he voluntarily decides to commit a murder, and this decision manifests very ill will. The revisionary Strawsonian could insist that when we focus on this fact, it is no longer intuitively obvious that Plum is not responsible in Pereboom's cases.

¹⁰ Some have even defended subjectivist accounts of wrongdoing, according to which wrong actions are *wrong* in that they reveal an objectionable quality of will. For a recent account along these lines, see Mason (2019).

¹¹ For more on the similarities between Frankfurt and Strawson, see McKenna (2004).

But in giving this kind of explanation behind the intuitive pull of the hard-line reply, we have run into trouble. In proceeding through Pereboom's cases as he suggests, we might realize that we *do* care about the origins of a person's abilities, including abilities of the sort needed to have and manifest quality of will. It matters to us that Plum, for instance, was manipulated by aliens or evil scientists poking around in his brain, even when Plum understands right from wrong and intentionally acts with ill will towards others. This is a fact about the very practices the revisionary Strawsonian also claims to be drawing on. Even if we can run the cases as the hard-liner suggests and reach ambiguity, we do sometimes have intuitions that are at odds with the revisionary, ahistorical approach described above. As such, we are unsure if the revised Strawsonian hard-line reply succeeds. As Pereboom notes, our practices require consistency. So, at the very least, the debate between the defender of the manipulation argument and the hard-liner should leave us ambivalent.

3.c. Response 2: Manipulation and determinism are relevantly different

Is there a way for the Strawsonian to accommodate the intuition that the causal source of an agent's abilities matters when it comes to moral responsibility without giving the game away to the incompatibilist? One way to do this would be to grant that the manipulative history of Plum in Cases 1–3 renders him not morally responsible, but to argue that there is a relevant difference between the manipulation featured in Cases 1–3 and determinism as its presented in Case 4. On this approach, the Strawsonian could object to the No Difference Premise, rather than deny the Manipulation Premise. In recent work, Andrew James Latham and Hannah Tierney (2021, 2022) have attempted to develop such a response to the four-case argument.

Latham and Tierney first make a distinction between universal and existential phenomena. Universal phenomena include every feature of the universe within their scope, while existential phenomena, in contrast, are those whose scopes include at least one but not all things within a universe (Latham and Tierney 2022: 150). In Cases 1–3, Plum is subject to existential manipulation—he is the only agent impacted by manipulation in these cases. However, Plum is impacted by determinism in Case 4, which is a universal phenomenon. If determinism is true, then, as Pereboom notes in his description of Case 4, “*Everything* that happens in our universe is causally determined by virtue of its past states together with the laws of nature” (Pereboom 2014: 79, emphasis added). Thus, Plum is the subject of existential manipulation in Cases 1–3, and simply one object among many within the scope of a universal phenomenon in Case 4.

Latham and Tierney take this to be a relevant difference, since existential and universal phenomena impact our practices in very different ways (2022: 151). In Cases 1–3, if we come to judge that Plum is not the fitting target of our moral responsibility practices because he was existentially manipulated, then our practices themselves are not threatened. While Plum may be exempt from our practices of praising, blaming, and holding responsible, no other agent will be exempted simply because Plum has. Latham and Tierney argue: “We can continue to participate in our other interpersonal relationships just as we did prior to learning about Plum and we can continue to praise, blame, and express the full range of reactive attitudes towards those with whom we are in these relationships” (2021: 151). But in Case 4, if we come to judge that Plum is exempt from our moral responsibility practices because his decision was causally determined, then we must exempt all other agents in our universe as well, since all agents' decisions (as well as everything else within the universe) is subject to the exact same universal phenomenon. But exempting all agents from our moral responsibility practices would destroy those practices and the relationships that ground them. And this is precisely what Strawson argues is

psychologically impossible to do when he makes his indispensability claim. So, while existential manipulation could rule out moral responsibility by functioning as an exemption within a set of practices in Cases 1–3, universal determinism can only rule out moral responsibility by eliminating our responsibility practices in Case 4. According to Latham and Tierney, this is a significant, and morally relevant, difference between Cases 1–3 and Case 4 and Strawsonians can reject the No-Difference Premise on these grounds (2021: 153).¹²

Unlike the hard-line response discussed above, this soft-line response is compatible with Strawson's original claim that unfortunate formative circumstances can exempt agents (1962: 66), at least when these circumstances are suitably novel. However, this view requires other modifications to the Strawsonian picture. In exploring what matters when it comes to our practices of praising, blaming, and holding others responsible, Strawson focuses almost exclusively on facts about agents' capacities, attitudes, and volitional states. But there is no difference between Plum's agential features in Cases 1–3 and Plum's agential features in Case 4—they are qualitatively identical across these cases. Rather, the relevant difference between Cases 1–3 and Case 4 involve *the circumstances* in which Plum's agential capacities and attitudes are exercised and expressed. Latham and Tierney argue:

If Plum is exempt in Cases 1–3, the underlying justification for this exemption rests not only on facts about Plum's agential capacities but also his circumstances, including the fact that he was the target of existential manipulation. This justification does not generalise to Plum in Case 4, since Plum is under the scope of a universal, not existential, phenomenon in this case. And one cannot generalise from existential exemptions to universal exemptions...the former operates within a practice while the later requires the destruction of it. (2021: 153–154)

So, in order to pursue the soft-line strategy outlined above, the Strawsonian will have to revise their view to account for the moral relevance of circumstantial factors to both our practices of praise and blame and to the nature of moral responsibility itself. This could require substantial, and controversial, revisions to the Strawsonian approach to moral responsibility.

It's also important to note that this response to Pereboom's manipulation argument relies on Strawson's indispensability claim, according to which exempting all agents from moral responsibility would require taking the objective stance towards others, thus destroying our ordinary interpersonal relationships and the practices of praise and blame that are built upon them. But this claim is controversial. Many, including Pereboom himself (2001, 2014, 2021), have argued that it is possible

¹² Latham and Tierney (2022) also consider two ways Pereboom could defend the No-Difference Premise: (1) replacing existential manipulation with universal manipulation in Cases 1–3, and (2) replacing universal determinism with existential determinism in Case 4. According to Latham and Tierney, (1) would place the Manipulation Premise in jeopardy, since it's unlikely that cases of universal manipulation can secure the intuition that Plum is not responsible (2022: 155). This claim is further supported by Latham and Tierney's empirical work on manipulation cases, where they found that participants judge universally manipulated agents to be significantly more free and responsible than existentially manipulated agents (2021: 7). Latham and Tierney go on to argue that while (2) could secure the Manipulation and No-Difference Premises, it cannot be used to defend the Anti-Compatibilist Conclusion (2022: 157). The Anti-Compatibilist Conclusion states that normally functioning agents in deterministic universes, because they are relevantly similar to Plum in Case 4, are not morally responsible for their actions. But normally functioning agents in deterministic universes are subject to universal determinism, and are thus not relevantly similar to Plum in a version of Case 4 where he is the subject of existential determinism. They argue: "...existential premises can only support existential conclusions, and the Anti-Compatibilist Conclusion is universal in nature" (2022: 157).

to maintain deep and meaningful relationships and forward-looking practices of praise and blame while also denying that agents can be morally responsible for their actions.¹³ So, this soft-line response relies on a contested feature of Strawson's original position. Can the Strawsonian defend this view? We examine this issue in the next section.

4. Are our praising and blaming practices really indispensable?

The soft-line reply considered in the last section requires a commitment to Strawson's idea that in the absence of praise and blame, we would not have ordinary interpersonal relationships as we know them. Universal exemptions would require us to view everyone from an objective point of view, thus destroying our moral practices and social lives. In response, the defender of the manipulation argument may appeal to the idea that we could have flourishing lives in the absence of emotional praise and blame, lives that would greatly preserve our ordinary relationships as we find them, and when revision requires diversion from these relationships, even improve them.

4.a. *A critique of angry blame*

One could argue that our current practices of praise and blame involve emotions that are themselves morally problematic and so better abandoned. The Strawsonian is committed to the idea that certain forms of anger are essential to blame, and so to a certain form of human relationship. Moral anger, anger at a perceived wrong or injustice, seems both fitting and useful. But one initial reason for hesitancy about the obvious good in anger stems from the diverse set of philosophical traditions that suggest we could do better without it. Consider some examples. The Buddha cautions that anger is vicious, and that we must control our bodies, speech, and minds so as not to let them be controlled by anger (Dhp XVII 231-233); Jesus of Nazareth says that morality doesn't just require that we judge the murderer; rather, the person who is angry is liable to judgement and the person who says "you fool" will be liable to hell (Matthew 5:21-23 NRSV); the Roman Stoic Seneca suggests that anger is "greedy for revenge" and "awkward at perceiving what is true and just" (De Ira, Book 1). According to this general line of criticism, anger is inherently bad because it is a distortion of real justice.

In what way is moral anger (of the sort the Strawsonian thinks constitutes blame) a distortion? Martha Nussbaum argues that anger conceptually involves a desire for retribution (2016: 15). According to Nussbaum, this retributive desire can be expressed in one of two ways: as a desire to socially down-rank (belittle) the target of one's anger or as a desire to get back at them for what they've done. But both of these desires are problematic—the desire to down-rank is objectionably narcissistic, while the desire for payback is irrational. Retribution might hurt the offender and cause them to suffer, but this cannot fulfill the desire for payback, since it will not do anything to erase the wrongdoing or make one's own suffering disappear. And retribution can succeed in downranking the offender but caring about one's social rank relative to others is petty. Why care about something so unimportant? Thus, on Nussbaum's view, moral anger cannot achieve justice.¹⁴

This critique of anger becomes even more powerful when one considers alternative attitudes that one can adopt in its place. Nussbaum (2016), for example, argues that problematic, backward-looking angry blame could "transition" into more positive forward-looking attitudes and responses. And Pereboom, who takes himself to be in broad agreement with Nussbaum, suggests that we could

¹³ See also Milam (2016).

¹⁴ For criticism of Nussbaum's view, see McBride (2018), Reis-Dennis (2018), and Srinivasan (2018).

replace angry blame with a forward-looking (and even confrontational) stance of *protest* against a wrongdoing (2021: 44-46).¹⁵ Protest allows us to stand against wrongdoing, while also fostering a sense of self-respect and dignity for victims, without exposing ourselves and others to the very real moral risks that tend to arise with the experience and expression of anger.

Even if there are alternatives to our current blaming practices, one might worry about what life will look like without it. But Pereboom (2014) paints an especially compelling picture. On his view, feelings of sadness, regret, joy, and certain forms of love can persist even if we judge that others are not the fitting or deserved targets of angry blame. According to Pereboom, eliminating angry, backward-looking blame from our lives is not just possible, it's practically and morally advisable. Doing so will not require us to maintain the objective stance towards everyone in our lives nor will it require the destruction of all *possible* praising and blaming practices, contrary to Strawson's claims. Rather, the elimination of angry blame will improve our interpersonal relationships and the praising and blaming practices that are built upon them.

4.b. *A defence of angry blame*

How might the Strawsonian respond to this challenge? One way of developing the Strawsonian view is to argue that there is an essential connection between our current, angry version of blame and the broader class of attitudes that make our lives meaningful and worth living. Seth Shabo (2012) has defended this contention by discussing the connection between blame as moral anger and love. Shabo argues that being in the sort of interpersonal relationships we value means *taking things personally*, which renders us susceptible to having our feelings hurt. And because we are the kinds of creatures that we contingently happen to be, hurt feelings will often give rise to species of moral anger like resentment (2012: 114). Going one step further, Anti Kauppinen (2018) has argued that, since getting angry is a way of valuing by caring, getting angry is connected to the whole emotional suite of valuing responses. What makes an emotional response a way of caring is a consistent, congruent pattern of (perhaps counterfactual) emotional evaluations. To care about something is not only to feel positive emotions in good cases, but negative emotions in bad cases, like when the object of care is threatened or put down. Indeed, this seems to be why anger is connected to self-respect,¹⁶ since getting angry over a slight to oneself is a display of self-care. In defending the moral importance of anger, the Strawsonian could also rely on recent work in analytic feminist and anti-racist philosophy, which suggests that there is an essential connection between anger and struggles for justice. For example, Myisha Cherry (2021) has argued that a certain type of anger, Lordean Rage (named after black feminist Audre Lorde and W.E.B. DuBois's term "black rage"), is uniquely good to support the anti-racist struggle. On Cherry's view, Lordean rage aims at social change by way of accountability for racist practices and the people who perpetuate them.¹⁷

¹⁵ Here is another example. In the context of clinical psychological settings, Hannah Pickard (2011, 2017) has argued that there can be forms of moral accountability in the absence of backward-looking desert. She suggests that we can distinguish between "affective blame", which involves the sting of negative emotions, and "detached blame", which can include forms of accountability without retributive anger. These forms of accountability might include the imposition of negative consequences or asking agents to account for their behaviour in the hope of raising moral awareness.

¹⁶ For other defenses of the relationship between angry blame and self-respect, see Dillon (1997); Murphy (2005); Reis-Dennis (2019); Tierney (2021).

¹⁷ One complication worth pursuing is whether Cherry's Lordean rage involves desert-entailing presuppositions, or the kind of fittingness that could be undermined by the truth of determinism. Is it possible that Lordean rage is relevantly similar to Pereboom's moral protest, just with a more overtly aggressive element? How should we think about our blaming responses to institutions and structures? Can we do so in terms of quality of will? Easy answers here are not obviously

At this stage, we've articulated two strikingly different views about the nature of angry blame. According to the anger sceptic, angry blame is a largely corrosive force in society and eliminating it from our lives will only stand to improve and enrich our relationships and practices. But according to the Strawsonian, angry blame plays an important role in the promotion of self-respect and justice. Eliminating it from our lives will only serve to numb us to caring about other people and morally important projects. What's gone wrong? How have these camps come to such radically different views of anger?

One possibility is that these theorists are not examining the same set of facts. While the Strawsonian is focused on our practices as they are, the sceptic is focused on what these practices could be. As Strawson notes, "Only by attending to this range of attitudes can we recover from the facts as we know them a sense of what we mean, i.e., of all we mean, when, speaking the language of morals, we speak of desert, responsibility, guilt, condemnation, and justice" (1962: 32). This strong sentimentalism highlights a difference in how the Strawsonian and sceptic understand justice. The sceptic of moral anger is trying to conceive of justice in the absence of the very attitudes the Strawsonian takes to exhaust the meaning of justice. In this sense, one might think that the moral anger sceptic is making a deep mistake.

However, the sceptic might respond that it is the Strawsonian who is mistaken when they fail to look beyond their own practices. As Per-Erik Milam (2017) argues, following Mill's line of thought in *The Subjection of Women*, it is extremely difficult to evaluate a normative practice *from within that very practice*. First, since none of us have ever seen a fully non-reactive society, we have nothing to compare our reactive society to. And second, any indirect evidence we get about piecemeal revisions to our current practices of praise and blame will be placed in a wider context where these attitudes have important and specific roles to play. A change that might ultimately be for the better might seem to cause local problems, making us falsely believe that the change is bad full stop. Compare with Mill's target audience who has never seen a society where women were not subjected, and who might balk at the idea of women holding office because they are partial while failing to realize that social structures restrict the sphere of women's influence to home and kin (2017: 6). There is a lot wrong with the reasoning in Mill's audience, but one important problem is that they fail to see how the wider context matters for evaluating a change. This point directly impacts arguments for reactive blame that appeal to the value of blame in contributing to social harmony (Nichols 2015; Vargas 2012). The point also should make us feel generally more sceptical about the kinds of evidence Strawsonians bring to bear on debates about the value of our current practices of praise and blame. Should we really trust the facts as we know them? This complication, at the very least, denies the Strawsonian the ability to take angry blame as the default position in the debate.

5. Conclusion

In this chapter we have focused on one kind of metaphysical threat to moral responsibility and how this threat could change current practices of praise and blame. But there are other metaphysical threats worth considering. Utilitarians might challenge the metaphysics of value inherent in a morality that demands just deserts even when this will bring about suffering without good. And philosophers

forthcoming. It remains to be seen if Strawsonians can absorb some of these alternatives into their view or if these complications favour the sceptics.

working in the tradition of Hume and Nāgārjuna might suggest that because there is no self as persisting substance, the concept of moral responsibility, along with the attending practices of praise and blame, are incoherent. In these metaphysical debates concerning value and personhood, we might see a similar dialectic play out between the sceptic and those who defend moral responsibility and our practices of praise and blame. Are praise and blame, as partially constituted or expressed by our emotions, a necessary and good feature of our humanity? Hume's own work offers a nice example of this tension between metaphysical threats and our common nature. He thinks we must allow ourselves a little anger, for "anger and hatred are passions inherent in our very frame and constitution" (Hume 1740: 605). Clearly, praise and blame are interestingly and inextricably linked to our moral psychology, our ethics, and our metaphysics. Thus, the relationship between moral responsibility, praise, and blame merits special attention.

Works Cited

- Arpaly, Nomy (2003). *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford University Press.
- Beglin, David (2018). Responsibility, libertarians, and the “facts as we know them”: A concern-based construal of Strawson’s reversal. *Ethics* 128: 612–625.
- Bennett, Jonathan (1980). Accountability. In Z. van Straaten (ed.) *Philosophical Subjects: Essays Presented to P.F. Strawson*. Oxford University Press.
- Brink, David & Nelkin, Dana (2013). Fairness and the architecture of responsibility. In D. Shoemaker (ed.) *Oxford Studies in Agency and Responsibility, volume 1*. Oxford University Press: 284–313.
- Caruso, Gregg D. (2021). *Rejecting Retributivism: Free Will, Punishment, and Criminal Justice*. Cambridge: Cambridge University Press.
- Cherry, Myisha (2021). *The Case for Rage: Why Anger is Essential to Anti-Racist Struggle*. New York: Oxford University Press.
- Ciurria, M. (forthcoming). Responsibility’s double binds: The reactive attitudes and conditions of oppression. *Journal of Applied Philosophy*.
- Chisholm, Roderick M., (1964). Human freedom and the self. In G. Watson (ed.) *Free Will*, 1st edition. New York: Oxford University Press: 24–35.
- Dillion, Robin (1997). Self-respect: moral, emotional, political. *Ethics* 107: 226–49.
- Fischer, John Martin & Ravizza, Mark (1998). *Responsibility and Control: An Essay on Moral Responsibility*. Cambridge: Cambridge University Press.
- Fischer, John Martin, Kane, Robert, Pereboom, Derk, & Vargas, Manuel (2007). *Four Views of Free Will*. Malden, MA: Blackwell Publishing.
- Frankfurt, Harry (1988) *The Importance of What We Care About: Philosophical Essays*. Cambridge: Cambridge University Press.
- Frankfurt, Harry (1971). Freedom of the will and the concept of a person. *The Journal of Philosophy*, 68(1): 5–20.
- Hume, David (1740/1978). *A Treatise of Human Nature*, P.H. Nidditch (ed.), Oxford: Clarendon Press.
- Kane, Robert (1996). *The Significance of Free Will*. New York: Oxford University Press.
- Kauppinen, Annti (2018). Valuing anger. In M. Cherry and O. Flanagan (eds.) *The Moral Psychology of Anger*. New York: Rowman and Littlefield.
- Latham, Andrew & Tierney, Hannah (2021). The four-case argument and the existential/universal effect. *Erkenntnis*: <https://doi.org/10.1007/s10670-021-00458-x>

- Latham, Andrew & Tierney, Hannah (2022). Defusing existential and universal threats to compatibilism: A Strawsonian dilemma for manipulation arguments. *Journal of Philosophy* 119: 144–161.
- Mason, Elinore (2019). *Ways to Be Blameworthy: Rightness, Wrongness, and Responsibility*. Oxford University Press.
- McBride, Lee (2018). Anger and approbation. In M. Cherry and O. Flanagan (eds.) *The Moral Psychology of Anger*. New York: Rowman and Littlefield.
- McKenna, Michael (2008). A hard-line reply to Pereboom’s four-case manipulation argument. *Philosophy and Phenomenological Research* 77(1): 142–159.
- McKenna, Michael (2012). Moral responsibility, manipulation arguments, and history: Assessing the resilience of nonhistorical compatibilism. *Journal of Ethics* 16: 145–174.
- Mele, Alfred (1995). *Autonomous Agents*. New York: Oxford University Press.
- Mele, Alfred (2006). *Free Will and Luck*. New York: Oxford University Press.
- Milam, Per-Erik (2016). Reactive attitudes and personal relationships. *Canadian Journal of Philosophy* 46: 102–122.
- Milam, Per-Erik (2017). In defence of non-reactive attitudes. *Philosophical explorations* 20(3): 294–307.
- Murphy, Jeffrie (2005). *Getting Even: Forgiveness and Its Limits*. Oxford University Press.
- Nelkin, Dana (2011). *Making Sense of Freedom and Responsibility*. Oxford University Press.
- Nichols, Shaun (2015). *Bound: Essays on Free Will and Responsibility*. Oxford University Press.
- Nussbaum, Martha (2016). *Anger and Forgiveness: Resentment, Generosity, Justice*. Oxford: Oxford University Press.
- Pereboom, Derk (2001). *Living Without Free Will*. Cambridge University Press.
- Pereboom, Derk (2014). *Free will, Agency, and Meaning in Life*. Oxford University Press.
- Pereboom, Derk (2017). Responsibility, regret, and protest. In D. Shoemaker (ed.) *Oxford Studies in Agency and Responsibility, volume 4*. Oxford University Press: 121–140.
- Pereboom, Derk (2021). *Wrongdoing and the Moral Emotions*. Oxford University Press.
- Pickard, Hanna (2011). Responsibility without blame: Empathy and the effective treatment of personality disorder. *Philosophy, Psychiatry, and Psychology* 18(3): 209–223.
- Pickard, Hanna (2017). Responsibility without blame for addiction. *Neuroethics* 10: 169–180.

- Reis-Dennis, Samuel (2019). Anger: Scary good. *Australasian Journal of Philosophy* 97: 451–64.
- Sartorio, Carolina (2016). *Causation & Free Will*. Oxford University Press.
- Strawson, Peter F. (1962). Freedom and resentment. In G. Watson (ed.), *Free Will*. New York, NY: Oxford University Press.
- Strawson, Peter F. (1985) *Skepticism and Naturalism: Some Varieties*. New York: Columbia University Press.
- Srinivasan, Amia. (2018) The aptness of anger. *The Journal of Political Philosophy* 26(2): 123–144.
- Tierney, Hannah (2021). Guilty confessions. In D. Shoemaker (ed.) *Oxford Studies in Agency and Responsibility, volume 7*. Oxford University Press: 182–204.
- van Inwagen, Peter (1983). *An Essay on Free Will*. Oxford: Clarendon Press.
- Vargas, Manuel. (2012). *Building better beings: A theory of moral responsibility*. Oxford University Press.
- Wallace, R. Jay (1994). *Responsibility and the Moral Sentiments*. Harvard University Press.
- Wolf, Susan (1980). Asymmetrical freedom. *The Journal of Philosophy* 77(3): 151–166.