**Defusing Existential and Universal Threats to Compatibilism: A Strawsonian Dilemma for Manipulation Arguments[[1]](#footnote-1)**

Compatibilist accounts of moral responsibility face a wide range of manipulation arguments.[[2]](#footnote-2) One popular variant of these manipulation arguments (roughly) takes the following form:

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.

Though there is an expansive literature dedicated to manipulation arguments, there is relatively little attention paid to the ways in which these arguments interact with Strawsonian approaches to compatibilism. In this paper, we argue that Strawsonian accounts of moral responsibility are immune to manipulation arguments and no attempt to modify the premises will help the incompatibilist secure their desired conclusion.

We first argue that manipulation, as it is described in many manipulation arguments, is nothing like determinism in one relevant respect. Determinism is a *universal* phenomenon: its scope includes every feature of the universe. However, manipulation arguments typically rely on cases in which only *one* agent is impacted by manipulation. We call manipulation whose scope includes at least one but not all agents *existential* *manipulation*. And universal phenomena operate very differently from existential phenomena. Even if our practices of holding others responsible are responsive to facts about existential manipulation, our practices will likely not be similarly responsive to a universal phenomenon like determinism. And this is a morally relevant difference between manipulation and determinism, especially on Strawsonian approaches to moral responsibility, which take facts about our responsibility practices to be importantly connected to the nature of moral responsibility itself. Thus, Strawsonians can reject the No-Difference Premise.

Next, we consider two ways of modifying manipulation arguments to resist our Strawsonian challenge. First, one could defend the No-Difference Premise by modifying it to feature universal manipulation instead of existential manipulation. Rather than considering cases in which only one agent is manipulated, one could present cases in which *all* agents are manipulated in the same way. We argue that while this could safeguard the No-Difference Premise, it would place the Manipulation Premise in serious jeopardy. Second, one could defend the No-Difference Premise by replacing universal determinism with existential determinism. That is, one could argue manipulation that targets a single agent is relevantly similar to determinism that impacts only one agent. We argue that while this move could secure the Manipulation and No-Difference Premises, it would come at the cost of the Anti-Compatibilist Conclusion, which is a claim about the incompatibility of universal, not existential, determinism and moral responsibility. In this way, Strawsonians can argue that manipulation arguments face a dilemma: either one of the premises is false or the premises cannot support the conclusion. After examining the implications of this argument, and clarifying the distinction between universal and existential phenomena, we conclude by exploring how this strategy can be adopted by non-Strawsonian compatibilists.

**I. Pereboom’s four-case manipulation argument**

In this paper, we focus on Derk Pereboom’s four-case manipulation argument,[[3]](#footnote-3) though we take our objection to generalise to all manipulation arguments that take the above form.[[4]](#footnote-4) We focus on this argument because it has received a considerable amount of attention from compatibilists and incompatibilists alike, and because it explicitly targets a wide range of compatibilist views, including Strawsonian accounts of moral responsibility.

In developing the four-case argument, Pereboom first asks his readers to consider 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 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.[[5]](#footnote-5)

Despite the fact that Plum meets a variety of sufficient compatibilist conditions for moral responsibility,[[6]](#footnote-6) Pereboom argues that intuitively he is not morally responsible for his decision to kill White.[[7]](#footnote-7) This indicates that these compatibilist accounts do not in fact feature the sufficient conditions for moral responsibility. Pereboom goes on to suggest that one explanation for the intuition that Plum is not morally responsible for his decision to kill White is that it was determined by factors (that is, the neuroscientists’ interventions) that were beyond Plum’s control.[[8]](#footnote-8) Next, Pereboom considers a case that is relevantly similar to Case 1 but slightly more ordinary:

Case 2: Plum is just like an ordinary human being, except that a team of neuroscientists programmed him at the beginning of his life so that his reasoning is often but not always egoistic (as in Case 1), and at times strongly so, with the intended consequence that in his current circumstances he is causally determined to engage in the egoistic reasons-responsive process of deliberation and to have the set of first and second-order desires that result in his decision to kill White. Plum has the general ability to regulate his actions by moral reasons, but in his circumstances, due to the strongly egoistic nature of his deliberative reasoning, he is causally determined to make his decision to kill. Yet he does not decide as he does because of an irresistible desire. The neural realization of his reasoning process and of his decision is exactly the same as it is in Case 1 (although their causal histories are different).[[9]](#footnote-9)

Although Plum again meets all the relevant sufficient compatibilist conditions on moral responsibility, Pereboom argues that Plum is intuitively not morally responsible for his decision to kill White. Furthermore, there are no relevant differences between Case 1 and Case 2, such that if you thought Plum was not responsible in Case 1, consistency requires that you form the same judgment about Case 2. Once Pereboom secures this intuition, he presents an even more ordinary case:

Case 3: Plum is an ordinary human being, except that the training practices of his community causally determined the nature of his deliberative reasoning processes so that they are frequently but not exclusively rationally egoistic (the resulting nature of his deliberative reasoning processes are exactly as they are in Cases 1 and 2). This training was completed before he developed the ability to prevent or alter these practices. Due to the aspect of his character produced by this training, in his present circumstances he is causally determined to engage in the strongly egoistic reasons-responsive process of deliberation and to have the first and second-order desires that issue in his decision to kill White. While Plum does have the general ability to regulate his behavior by moral reasons, in virtue of this aspect of his character and his circumstances he is causally determined to make his immoral decision, although he does not decide as he does due to an irresistible desire. The neural realization of his deliberative reasoning process and of the decision is just as it is in Cases 1 and 2.[[10]](#footnote-10)

Pereboom contends that there are no relevant differences between Plum in Case 3 and Plum in Cases 1 and 2. Therefore, if we judge that Plum is not morally responsible in Cases 1 and 2, then we should also judge that Plum is not responsible in Case 3. Lastly, Pereboom presents the fourth and final case:

Case 4: Everything that happens in our universe is causally determined by virtue of its past states together with the laws of nature. Plum is an ordinary human being, raised in normal circumstances, and again his reasoning processes are frequently but not exclusively egoistic, and sometimes strongly so (as in Cases 1–3). His decision to kill White issues from his strongly egoistic but reasons-responsive process of deliberation, and he has the specified first and second-order desires. The neural realization of Plum’s reasoning process and decision is exactly as it is in Cases 1–3; he has the general ability to grasp, apply, and regulate his actions by moral reasons, and it is not because of an irresistible desire that he decides to kill.[[11]](#footnote-11)

Once again, Pereboom argues that Plum in Case 4 is relevantly similar to Plum in Cases 1, 2, and 3, thus consistency compels us to judge that Plum is not morally responsible for his decision to kill White in Case 4. If this is right, then agents who are like Plum in Case 4, that is, agents who live in deterministic worlds, also fail to be morally responsible. In other words, compatibilism about moral responsibility is false.

We can now flesh out the manipulation argument schema with which we began the paper with details from Pereboom’s four-case argument.

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.

There is much to say in favour of Pereboom’s careful and forceful argument. Notably, this argument challenges a host of compatibilist accounts of moral responsibility, including Strawsonian approaches as well. In “Freedom and Resentment,” P. F. Strawson argues that the truth of determinism is irrelevant to our practices of praising, blaming, and holding others morally responsible.[[12]](#footnote-12) While these practices are responsive to facts about agents’ intentions, beliefs, and cognitive capacities, learning that determinism is true will do nothing to diminish the degree to which we praise, blame, and hold others morally responsible. And, for Strawson, being the appropriate target of these practices, particularly the reactive attitudes, constitutes *being* morally responsible. So, on Strawson’s view, determinism is not only irrelevant to these practices and attitudes, it is irrelevant to moral responsibility itself.

However, Pereboom argues that the four-case argument illustrates that determinism *is* relevant to our practices of holding others morally responsible. As Cases 1–3 show, our practices are responsive to facts about manipulation. When we learn that Plum was manipulated to decide to kill White, we judge that he is not morally responsible for this decision. And, Pereboom contends, our responsibility practices are sensitive to considerations of arbitrariness: “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.”[[13]](#footnote-13) Because Plum in Case 4 is relevantly similar to Plum in Cases 1–3, our responsibility practices dictate that we should also exempt Plum in Case 4 from blame. If this is right, then determinism is in fact relevant to our responsibility practices and, given Strawson’s view, moral responsibility as well.

**II. Existential manipulation and universal determinism**

Though much has been written about the four-case argument, relatively little attention has been paid to Pereboom’s discussion of Strawsonian compatibilism. We attempt to remedy this by exploring a Strawsonian objection to the four-case argument.

In its current form, Strawsonians should reject the No-Difference Premise of the four-case argument. According to this premise, Plum in Case 4 is not morally responsible for his decision to kill White because he is relevantly similar to Plum in Cases 1–3. While Pereboom may be right that *Plum* possesses all of the same compatibilist-friendly capacities in Cases 1–4, facts about Plum are not the only morally relevant features of these cases. There is a significant difference between the circumstances in which Plum finds himself in Cases 1–3 and Case 4, and this difference in circumstance is relevant to our practices of holding, and thus being, responsible on a Strawsonian view of moral responsibility.

In Cases 1–3, Plum is the target of *existential* manipulation—he is the onlyagent impacted by manipulation in these cases. Indeed, the first sentence of both Case 2 and Case 3 indicates that Plum is different from ordinary human beings *because* he is manipulated.[[14]](#footnote-14) In contrast, Plum is not the only agent impacted by determinism in Case 4. Determinism, as it is presented in Case 4, is a *universal* phenomenon—its scope includes everything within our universe. We are told that “*Everything* that happens in our universe is causally determined by virtue of its past states together with the laws of nature.”[[15]](#footnote-15) So, while Plum is the single target of existential manipulation in Cases 1–3, he is simply one object among many within the scope of a universal phenomenon in Case 4.

It is possible that, as Pereboom argues, our practices are responsive to facts about existential manipulation. When we come to learn that Plum’s decisions are determined by a team of neuroscientists or the training practices of his community, we may judge that he is exempt from being morally responsible.[[16]](#footnote-16) That is, we may take what Strawson calls the objective stance towards Plum and see him as an individual who cannot engage in the kinds of relationships that make the range of reactive attitudes fitting or appropriate. But even if we grant that our moral responsibility practices are responsive to existential manipulation, we need not grant that these practices are *also* responsive to universal determinism.

One morally relevant difference between existential manipulation and universal determinism is how these phenomena could come to affect our responsibility practices. In Cases 1–3, if we come to judge that Plum, because he was the target of existential manipulation, is not the fitting target of our moral responsibility practices, our practices themselves will not be threatened. Though Plum is the victim of existential manipulation, no one else in our universe is. So, while Plum may be exempt from our responsibility practices and we may come to take the objective stance towards him, no other agent in our universe will suffer the same fate simply because Plum has. 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.

But this is not so when it comes to Case 4. In this case, we learn that universal determinism is true and that Plum’s decisions, like every agents’, are causally determined by facts about the past in conjunction with the laws of nature. If we come to judge that Plum is not the fitting target of our moral responsibility practices because his decision was causally determined, then we must also make the same judgment about every other agent in our universe as well.[[17]](#footnote-17) For Plum is not the only agent whose decisions are causally determined if universal determinism is true. Rather, all agents will be in precisely the same circumstances as Plum. And, as Pereboom argues: “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.”[[18]](#footnote-18) So, in taking the objective stance towards Plum, we must also commit to taking the objective stance towards every other agent in our universe, including ourselves. And according to Strawson, universally viewing agents from the objective stance would not just threaten our responsibility practices, it would destroy them. If we were to take the objective stance towards everyone in our lives, then we could not maintain any of our ordinary interpersonal relationships or express the wide range of reactive attitudes made appropriate by these relationships. This is a significant cost of judging that Plum is not morally responsible in Case 4. Indeed, it is a cost that Strawson argues would be impossible, both practically and morally, to bear:[[19]](#footnote-19)

The human commitment to participation in ordinary interpersonal relationships is, I think too thoroughgoing and deeply rooted for us to take seriously the thought that a general conviction might so change our world that, in it there were no longer any such things as interpersonal relationships as we normally understand them; and being involved in interpersonal relationships as we normally understand them precisely is being exposed to the range of reactive attitudes and feelings in question.[[20]](#footnote-20)

On Strawson’s view, taking the truth of universal determinism to exculpate Plum in Case 4 would require an impossible feat: the elimination of our responsibility practices and the relationships that these practices are built upon.

 While Strawson’s claims have garnered support in the literature,[[21]](#footnote-21) they have also been challenged,[[22]](#footnote-22) most notably by Pereboom himself.[[23]](#footnote-23) Pereboom argues that while free will scepticism undermines basic desert responsibility and renders certain reactive attitudes irrational, it is compatible with forward-looking responsibility and a range of attitudes and emotions that are sufficient to sustain meaningful personal relationships. Still, Pereboom is clear that free will scepticism does require a change to both our blaming practices and interpersonal relationships, and much of his recent work is dedicated to clarifying the nature of these revisions.[[24]](#footnote-24) Thus, while there is a debate to be had about how much revision to our practices and relationships is required by universally denying basic desert responsibility, the fact that they must be revised under such circumstances is not contested. So, even if exculpating Plum in Case 4 does not require the elimination of our interpersonal relationships, it will require us to revise these relationships, along with our responsibility practices. And this is all the Strawsonian needs to reject the No-Difference Premise and argue that there is a relevant difference between Plum in Cases 1–3 and Plum in Case 4.

 In Cases 1–3, manipulation rules out Plum’s (basic desert) responsibility by functioning as an exemption within a set of practices. The reactive attitudes, our interpersonal relationships, and our responsibility practices can all go untouched if we take existential manipulation to exempt Plum. But in Case 4, if determinism rules out Plum’s (basic desert) responsibility, it would not do this via functioning as an exemption within a practice, but rather by requiring that the practices themselves be destroyed or altered. And this is a significant, and morally relevant, difference between Cases 1–3 and Case 4. Notice that this is so even though there is no morally relevant difference between *Plum’s agential capacities* across the cases. Agents’ capacities are not the only features that are morally relevant when it comes to understanding the nature of moral responsibility within a Strawsonian framework. The circumstances in which these capacities are exercised is also morally relevant. This is because our relationships, and the praising and blaming practices that are scaffolded around them, are sensitive to agential and circumstantial facts. So, 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 for exemption 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 in a Strawsonian framework—the former operates within a practice while the later requires the destruction or alteration of it.

**III. The Strawsonian Dilemma**

Identifying a relevant difference between Cases 1–3 and Case 4 may not keep the incompatibilist at bay for long. Pereboom could overcome the objection laid out in the above section by either modifying the scope of manipulation featured in Cases 1–3 or the scope of determinism featured in Case 4. This is precisely why Michael Mckenna warns compatibilists against merely rejecting the No-Difference Premise of manipulation arguments.[[25]](#footnote-25) However, we will argue that while modifying the scope of manipulation in Cases 1–3 or the scope of determinism in Case 4 will allow Pereboom to defend the No-Difference Premise of the manipulation argument, these strategies will ultimately be unsuccessful.

 *III.1. Horn One: Universal Manipulation and Universal Determinism*

Pereboom could replace the instances of existential manipulation in Cases 1–3 with universal manipulation. In Case 1 we would be asked to imagine that a team of neuroscientists have the ability to manipulate all human beings’ neural states, including Plum’s, at any time by radio-like technology. It is important to note that because this is a case of universal manipulation, the team of neuroscientists would also be subject to the same intervention, as well as those who manipulate them, and so on and so forth. In Case 2 we would be asked to imagine that all human beings, including Plum, are programmed by a team of neuroscientists at the beginning of their lives. And, like in Case 1, it is important to remember that this includes the team of neuroscientists (and those who programmed these neuroscientists, etc.) as well. Finally, in Case 3, we would be asked to imagine that all ordinary human beings, including Plum, are subject to the training practices of their communities.

If Cases 1–3 featured universal manipulation, then they would arguably be relevantly similar to Case 4. In all four cases, Plum would have the exact same agential capacities *and* he would find himself in relevantly similar circumstances. Thus, such modifications could allow Pereboom to defend the No-Difference Premise of the four-case argument. However, these modifications would also place the Manipulation Premise in serious jeopardy. It is not at all clear that these cases would elicit the intuition that Plum is not morally responsible for his decision to kill White because of the way in which he was manipulated. Furthermore, the Strawsonian can reject this claim using the same strategy we outlined above to object to the claim that Plum is not morally responsible in Case 4.

In the universal manipulation formulations of Cases 1–3, every other agent in the universe is the target of the exact same kind of manipulation, including those who manipulated Plum. So, if Plum is not responsible in these cases in virtue of being manipulated in the way that he was, then no one is. And this would mean that the practices of holding others morally responsible, and the relationships that make these practices appropriate and fitting, would be destroyed or significantly altered. But, as Strawson argues, “the human commitment to participation in ordinary interpersonal relationships is…too thoroughgoing and deeply rooted” to take this possibility seriously.[[26]](#footnote-26) So, it cannot be the case that Plum is not morally responsible for his decision in virtue of being manipulated in the way that he was. Thus, the Strawsonian can reject the Manipulation Premise.

While Pereboom may be right that *existential* “manipulation-fueled exemptions to moral responsibility” are widely accepted,[[27]](#footnote-27) it does not follow from this that *universal* manipulation has the same status. Though existential manipulation exemptions do not threaten our interpersonal relationships or responsibility practices, taking universal manipulation to exculpate Plum very much would. And it is our commitment to these relationships and practices that renders universal manipulation incapable of exculpating Plum. Just as Pereboom’s original generalisation strategy fails when moving from existential to universal cases, so too does the strategy of replacing existential cases with universal ones.

 *III.2. Horn Two: Existential manipulation and existential determinism*

There is another way Pereboom can defend the No-Difference Premise. Rather than replace existential manipulation with universal manipulation in Cases 1–3, he could modify the determinism featured in Case 4 to be existential rather than universal. This would involve asking readers to imagine that Plum is an ordinary human being, except that he is unique in that everything he does is causally determined by virtue of past states together with the laws of nature.[[28]](#footnote-28) With this modification, one could grant the No-Difference Premise and agree that Cases 1–3 and Case 4 are relevantly similar: Plum, in each case, possesses the same agential capacities and finds himself in relevantly similar circumstances. But doesn’t this put the compatibilist in a difficult position? If one also grants the Manipulation Premise, and the existence of existential manipulation exemptions, doesn’t this mean that one is committed to the claim that Plum is not morally responsible in the existential formulation of Case 4 as well?

Simply put: Yes.[[29]](#footnote-29) And perhaps this outcome is intuitive. If Plum really is different from every other agent in the universe in virtue of his behaviour being determined, then there may be a sense in which Plum has less control over his actions than ordinary agents.[[30]](#footnote-30) And it is possible that the practices of praising, blaming, and holding others responsible would be sensitive to this kind of difference. If we come to know that a single individual in our universe is subject to determinism while every other agent is not, this may indeed affect our reactions to this individual and the relationships we cultivate with them. But if one grants that Plum in the existential determinism formulation of Case 4 is not morally responsible, does this mean that one is also committed to the Anti-Compatibilist Conclusion?

Not at all. 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 *not* relevantly similar to Plum in the existential determinism formulation of Case 4. Agents who live in deterministic universes are subject to universal determinism. But in the modified version of Case 4, Plum, and only Plum, is subject to existential determinism. A Strawsonian compatibilist can grant that agents who are existentially determined are not responsible without committing themselves to the claim that agents who are subject to universal determinism suffer the same fate. In fact, they have a principled reason to reject such an implication. While our blaming practices can be responsive to existential theses, they are not (and cannot be, on threat of destruction or revision) responsive to universal theses. Thus, though modifying Case 4 to feature existential determinism can allow Pereboom to defend the Manipulation Premise and the No-Difference Premise, doing so will do nothing to help him defend the Anti-Compatibilist Conclusion. In short, existential premises can only support existential conclusions, and the Anti-Compatibilist Conclusion is universal in nature.

There are a few interesting implications of this line of argumentation. First, on the above approach, compatibilists can grant that certain kinds of determinism undermine moral responsibility. Thus, there is one sense in which Pereboom’s criticism of Strawson is correct. Pereboom argues that the four-case argument “…allows us to see that the challenge determinism poses to moral responsibility and the associated reactive attitudes is not based on justificatory requirements external to the practice of holding people morally responsible.”[[31]](#footnote-31) And this may very well be true of *existential* determinism. But in another sense this criticism is misguided. *Universal* determinism, the phenomenon in question in debates over compatibilism and incompatibilism, holds, almost by definition, outside of our responsibility practices. If we were to try to incorporate universal determinism exemptions into our practices, they would fall apart or require significant revision, as would the interpersonal relationships that make these practices appropriate.

One might think that conceding that existential determinism can undermine moral responsibility will lead the compatibilist down a slippery slope whereby they will be forced to ultimately accept that universal determinism undermines backward-looking, desert-invoking moral responsibility as well. But the Strawsonian compatibilist is well-placed to make an in-principle distinction between these two varieties of determinism. Existential determinism exemptions are compatible with the continuation of our responsibility attributing practices, but universal determinism exemptions are not. Because we cannot realistically give up or significantly alter these practices, the Strawsonian compatibilist can deny that universal determinism is relevant to our moral responsibility practices, and thus moral responsibility itself, while nonetheless granting that existential determinism exemptions may very well be a feature of our practices. As we have argued, generalising from existential manipulation to universal determinism does not succeed against the Strawsonian. Likewise, the strategy of generalising from existential determinism to universal determinism will also fail.

In sum, both ways of securing the No-Difference Premise of the four-case argument will not be successful. Replacing the existential manipulation of Cases 1–3 with universal manipulation will render these cases relevantly similar to Case 4, but this modification puts the Strawsonian in the position to reject the Manipulation Premise. Similarly, one can defend the No-Difference premise by altering Case 4 to feature existential determinism rather than universal determinism, but this creates a serious disconnect between the premises of the four-case argument and its Anti-Compatibilist Conclusion.

**IV. Objection: Existential versus Universal Manipulations**

While a universal phenomenon’s scope includes every agent, an existential phenomenon affects *at least* one but not all agents. However, in this paper we have considered instantiations of existential manipulation and determinism that impact only one agent. What about existential manipulation that affects 99.9 percent of all agents? According to our proposal, our practices can be responsive to existential manipulation exemptions. But in a case where 99.9 percent of agents are exempt from our praising, blaming, and responsibility attributing practices, could these practices survive or would they be destroyed just as they would be if we were to incorporate universal manipulation exemptions into them?

 To be clear, our claim is not that our practices can incorporate *all* existential exemptions. We only claim that our practices can incorporate *some* of these exemptions, like those featured in Pereboom’s original Cases 1–3 and the existential version of Case 4. We think it is most likely that our practices would be destroyed if we were to try to incorporate exemptions that would apply to 99.9 percent of agents. This is because ordinary agents and their responsibility practices are not sensitive to subtle differences. Just as we should expect to be unable to distinguish between a phenomenon that impacts 99.8 percent of agents and a phenomenon that impacts 99.9 percent of agents, we should also expect to be unable to distinguish between a phenomenon that affects 99.9 percent of all agents from one that affects 100 percent of those agents. This means that we will likely end up treating any exemption which applies to 99.9 percent of agents as if it impacted all agents, and so such exemptions should not, and cannot, be incorporated into our responsibility practices at all. Further, even *if* we could make this distinction, practices that hold only .1% of agents responsible would be next to useless in playing the roles those practices are ordinarily meant to play.

With this in mind, one might wonder where the threshold between an existential manipulation and a universal manipulation should be set. We do not have an answer to this question, or any firm intuitions about where this line should be drawn. The fact that we do not have any solid intuitions about this matter should not be surprising. Remember that moral responsibility on the Strawsonian approach is a form of realism with respect to our ordinary responsibility practices. While we should expect to have solid intuitions regarding clear cut cases where either one agent (or perhaps a local niche population) or all agents are subject to some manipulation, we should also expect to have weak, and possibly confused, intuitions about borderline cases. After all, it is simply not clear what our ordinary responsibility practices would be like, or should be expected to be like, in such cases.

To illustrate, consider the following case. Imagine a world that consists of two populations of equal numbers: Population A and Population B. While the agents of Population A are all manipulated (perhaps just as Plum is manipulated in one of Pereboom’s original cases), the agents of Population B are not. Imagine, further, that no one is aware of the role manipulation plays in their world and all agents deploy our ordinary responsibility practices.

There are a number of ways the responsibility practices of such a world could evolve once the fact that agents of Population A are all manipulated comes to light.[[32]](#footnote-32) It is possible to imagine a scenario in which Population B discovers that Population A are manipulated, but also discovers their ordinary responsibility practices towards them are intact. As a result, Population B comes to judge that this kind of manipulation is irrelevant to responsibility practices and moral responsibility itself. In this scenario the fact that 50 percent of all agents’ actions are the product of this kind of manipulation turns out to have no effect on the blaming practices of this world. But it is also possible to imagine a scenario in which Population B comes to judge that Population A are not morally responsible and so exempts these agents from the responsibility practices. We can also imagine Population A continuing to target other members of Population A (and perhaps even Population B) with ordinary responsibility practices, even though members of Population B do not target members of Population A. Ultimately then, our blaming practices in borderline cases depend on many factors, and the conceptual relationship between existential manipulation and moral responsibility is only *one* of these features.

Concerns regarding how existential manipulation affect our blaming practices in borderline cases, and where the line should be drawn between existential and universal manipulation, are unlikely to assist the incompatibilist in their argument against the compatibilist. The considerations above do not indicate that there is something particularly responsibility undermining about manipulation or determinism itself. Our practices are entirely compatible with universal instantiations of both manipulation and determinism after all, contra the incompatibilist. Rather, the way manipulation and determinism can affect our responsibility practices, and responsibility itself, depends, at least partially, on the *scope* of the phenomena. But the incompatibilist’s arguments do not hang on how the scope of a phenomena affects responsibility. Interestingly, most compatibilists are similarly silent with respect to issues of scope in their accounts of moral responsibility. In the next, and final, section of this paper, we suggest that compatibilists would do well to attend to this lacuna.

**V. Conclusion: Beyond Strawson**

In this paper, we have argued that, contra Pereboom, the four-case argument does not undermine Strawsonian accounts of moral responsibility. This is because there is a significant and relevant difference between existential phenomena and universal phenomena. While our ordinary responsibility practices might be responsive to existential manipulation and existential determinism, universal manipulation and universal determinism cannot impact these practices. And, because Strawsonian accounts of moral responsibility take being *held* responsible to constitute *being* responsible, universal determinism cannot affect whether an agent ismorally responsible.

But what about those who reject Strawsonian approaches to moral responsibility? While Strawson’s work on determinism and moral responsibility is hugely influential, what is often called the “Strawsonian reversal” is quite controversial and some take it to be implausibly undemanding.[[33]](#footnote-33) Many have interpreted Strawson as arguing that an agent’s being morally responsible depends on it being appropriate to hold that agent morally responsible.[[34]](#footnote-34) And while many agree that our moral responsibility practices are important, few defend the claim that being the target of such practices constitutes being responsible*.*[[35]](#footnote-35)Rather, many compatibilists take an instance of praise, blame, or some other attribution of responsibility to be justified by some further fact, typically about the agent in question. This means that simply because our moral responsibility *practices* are immune to the threat of universal determinism, it does not automatically follow that moral responsibility itself is also immune on non-Strawsonian accounts.

Can our proposal safeguard othercompatibilist accounts from the four-case manipulation argument? While those who reject the “Strawsonian reversal” will not be able to adopt the *exact* strategy we outline in this paper, they can still make use of this approach in developing responses to the four-case argument and, perhaps more importantly, in improving their own accounts of moral responsibility. Though many compatibilists do not take our responsibility practices to constitute responsibility itself, they still take these practices to be important. And they care about these practices because they take them to (defeasibly) track facts that are relevant to responsibility. And while compatibilists, and those who work on moral responsibility more generally, focus almost exclusively on intrapersonal facts about agents to inform their views, one point we have stressed is that an agent’s capacities are not the only features which are relevant when it comes to our responsibility practices. Facts about the circumstances in which these capacities are exercised also matter. Once non-Strawsonian compatibilists acknowledge these interpersonal facets of our practices, they should attempt to capture them in their theories. Once they do, they too can adopt our strategy for dealing with Pereboom’s four-case argument and other manipulation arguments.

1. The authors contributed equally to this work. Many thanks to Michael McKenna, Derk Pereboom, Robert H. Wallace, David Braddon-Mitchell, Kristie Miller, the DaGERS reading group, the University of Sydney metaphysics reading group, and two anonymous referees for this journal for very helpful feedback on earlier drafts of this essay. Andrew J. Latham would also like to thank Ngāi Tai Ki Tāmaki Tribal Trust for their support. [↑](#footnote-ref-1)
2. See Robert Kane, *The Significance of Free Will* (New York: Oxford University Press, 1996); Alfred Mele, *Autonomous Agents* (New York: Oxford University Press, 1995); Alfred Mele, *Free Will and Luck* (New York: Oxford University Press, 2006); Derk Pereboom, *Living Without Free Will* (Cambridge: Cambridge University Press, 2001); Derk Pereboom, *Free Will, Agency, and Meaning in Life* (Oxford: Oxford University Press, 2014). [↑](#footnote-ref-2)
3. Pereboom, *Living Without Free Will, op. cit*.; Pereboom, *Free Will, Agency, and Meaning in Life, op. cit*. [↑](#footnote-ref-3)
4. For example, our objection applies equally well to Mele’s (*Free Will and Luck, op. cit.*) zygote argument. Mele first presents a case in which a powerful goddess intentionally creates a zygote within a deterministic universe in order to ensure that this being will develop into an agent who performs a particular action thirty years later. Mele first suggests that such an agent is intuitively not morally responsible because of the way in which his zygote was produced (importantly, Mele himself is agnostic about this claim). He then argues that there is no relevant difference between the way this agent’s zygote was produced and the way normal human zygotes come to be in deterministic universes, then determinism is incompatible with moral responsibility (*Ibid*., p. 189). Because Mele’s manipulation case features an instance of existential manipulation, where a single agent’s zygote was intentionally created by a powerful goddess, and he argues that there is no relevant difference between this case and a case in which determinism, a universal phenomenon, occurs, it is vulnerable to the critique we develop in this paper. [↑](#footnote-ref-4)
5. Pereboom, *Free Will, Agency, and Meaning in Life, op. cit*., pp. 76–77. [↑](#footnote-ref-5)
6. For compatibilist accounts of the sufficient conditions for moral responsibility, see David Hume *A Treatise of Human Nature* (1739); Harry Frankfurt, “Freedom of the Will and the Concept of a Person,” this journal, lxviii, 5 (1971): 5–20; J. M. Fischer and M. Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* (New York: Cambridge University Press, 1998); and R. Jay Wallace, *Responsibility and the Moral Sentiments* (Cambridge, MA: Harvard University Press, 1994). [↑](#footnote-ref-6)
7. One might argue that Plum does not meet the relevant conditions for responsibility or that the kind of manipulation described in this case is impossible. Indeed, several theorists have raised questions about the plausibility of Case 1 (J. M. Fischer, “Responsibility and Manipulation,” *The Journal of Ethics,* viii, 2 (2004): 145–177;Michael McKenna, “A Hard-Line Reply to Pereboom’s Four-Case Manipulation Argument,” *Philosophy and Phenomenological Research,* lxxvii, 1 (2008): 142–159; Alfred Mele, “A Critique of Pereboom’s ‘Four-Case’ Argument for Incompatibilism,” *Analysis,* lxv (2005): 75–81). However, many ultimately go on to grant that it is possible (Fischer, “Responsibility and Manipulation,” *op. cit.*) or present charitable reconstructions of the case (McKenna, “A Hard-Line Reply to Pereboom’s Four-Case Manipulation Argument,” *op. cit.*). Because we are interested in exploring how much the compatibilist can grant to defenders of manipulation arguments while still denying the Anti-Compatibilist Conclusion, we will grant Pereboom’s description of his cases. [↑](#footnote-ref-7)
8. Pereboom, *Free Will, Agency, and Meaning in Life, op. cit*., p. 77. [↑](#footnote-ref-8)
9. *Ibid*., p. 77. [↑](#footnote-ref-9)
10. *Ibid*., p. 78. [↑](#footnote-ref-10)
11. *Ibid.,* p. 79. [↑](#footnote-ref-11)
12. P.F. Strawson, “Freedom and Resentment,” in Gary Watson, ed., *Proceedings of the British Academy, volume 48: 1962* (Oxford: Oxford University Press, 1962). [↑](#footnote-ref-12)
13. Pereboom, *Free Will, Agency, and Meaning in Life, op. cit*., p. 82. [↑](#footnote-ref-13)
14. *Ibid*., pp. 77–78. [↑](#footnote-ref-14)
15. *Ibid*., pp. 79, emphasis added. [↑](#footnote-ref-15)
16. A Strawsonian compatibilist need not grant this premise. Rather, they could take what McKenna calls the hard-line and deny this claim (“A Hard-Line Reply to Pereboom’s Four-Case Manipulation Argument,” *op. cit.*; “Resisting the Manipulation Argument: A Hard-Liner Takes it on The Chin,” *Philosophy and Phenomenological Research,* lxxxix, 2 (2014): 464–484). However, we are interested in exploring how much the compatibilist can grant to the defender of the four-case argument while still denying its Anti-Compatibilist Conclusion. [↑](#footnote-ref-16)
17. One might argue that Pereboom’s cases are meant to be about worlds distinct from our own, such that we could judge that Plum, and every member of *his* world are exempt from being responsible, but still maintain our own blaming practices. But notice that this would only get the incompatibilist the Anti-Compatibilist Conclusion about this hypothetical worldand not *our* world, which is surely what they are after. Further, as we argue in section III.2., it will be no small feat to move from the claim that determinism rules out moral responsibility in *some* world to *our* world. See footnote 19 for empirical work that suggests that ordinary people are sensitive to this distinction. [↑](#footnote-ref-17)
18. Pereboom, *Free Will, Agency, and Meaning in Life, op. cit*., p. 82. [↑](#footnote-ref-18)
19. Strawson’s prediction has largely been born out in the research. For example, see Adam Feltz, “Pereboom and Premises: Asking the Right Questions in the Experimental Philosophy of Free Will,” *Consciousness and Cognition*, xxii, 1 (2013): 53–63; Andrew Latham, “The Conceptual Impossibility of Free Will Error Theory,” *European Journal of Analytic Philosophy*, xv, 2 (2019): 99–120; Eddy Nahmias, Stephen Morris, Thomas Naddelhoffer, and Jason Turner, “Surveying Freedom: Folk Intuitions about Free Will and Responsibility,” *Philosophical Psychology,* xviii, 5 (2005): 561–584; Adina Roskies and Shaun Nichols, “Bringing Moral Responsibility Down to Earth,” this Journal, cv, 7 (2008): 371–388; Though for empirical work that calls Strawson’s predication into question, see Shaun Nichols, “The Folk Psychology of Free Will: Fits and Starts,” *Mind & Language,* xix, 5 (2004): 473–502; Shaun Nichols and Joshua Knobe, “Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions,” *Noûs,* xli, 4 (2007): 663–685; Hagop Sarkissian, Amita Chatterjee, Filipe De Brigard, Joshua Knobe, Shaun Nichols, and Smita Sirker, “Is Belief in Free Will a Cultural Universal?” *Mind & Language* xxv, 3 (2010): 346–358.

 [↑](#footnote-ref-19)
20. Strawson, “Freedom and Resentment,” *op. cit*., p. 66. [↑](#footnote-ref-20)
21. See, for example, Seth Shabo, “Where Love and Resentment Meet: Strawson’s Intrapersonal Defence of Compatibilism,” *Philosophical Review*, cxxi (2012): 95–124; Susan Wolf, “The Importance of Free Will,” *Mind,* ccclix (1981): 386–406. [↑](#footnote-ref-21)
22. Pere-Erik Milam “Reactive Attitudes and Personal Relationships,” *Canadian Journal of Philosophy*, xlvi (2016): 102 122. [↑](#footnote-ref-22)
23. Pereboom, *Living Without Free Will, op. cit*.; Pereboom, *Agency, and Meaning in Life, op. cit*. [↑](#footnote-ref-23)
24. Derk Pereboom, “Responsibility, Regret, and Protest,” in David Shoemaker, ed., *Oxford Studies*

*in Agency and Responsibility*, *volume 4* (Oxford: Oxford University Press, 2017), pp. 121–140; Derk Pereboom, “A Forward-Looking Account of Self-Blame” in Andreas Carlsson, ed., *Self-Blame and Moral Responsibility* (Cambridge: Cambridge University Press). [↑](#footnote-ref-24)
25. McKenna, “A Hard-Line Reply to Pereboom’s Four-Case Manipulation Argument,” *op. cit.*; McKenna, “Resisting the Manipulation Argument: A Hard-Liner Takes it on The Chin,” *op. cit.*  [↑](#footnote-ref-25)
26. Strawson, “Freedom and Resentment,” *op. cit.,* p. 66. For an argument that our commitment to our practices and relationships is too thoroughgoing to even allow for revision, as opposed to destruction, see David Shoemaker, “Response-Dependent Responsibility; Or, a Funny Thing Happened on the Way to Blame Response-Dependent Responsibility,” *Philosophical Review* cxxvi, 4(2017): 481–527, at pp. 513–516. Of course, Pereboom would likely not be convinced by this argument. After all, his recent work argues that free will scepticism will require relatively minimal changes to our practices and relationships (Pereboom, “Responsibility, Regret, and Protest,” *op. cit*.; Pereboom, “A Forward-Looking Account of Self-Blame,” *op. cit.*), and he could make similar arguments with regards to universal manipulation. But notice that this potential defence of the Manipulation Premise relies on the success of independent arguments for the viability of free will scepticism, which would render the manipulation argument incapable of independently challenging compatibilist accounts of moral responsibility. [↑](#footnote-ref-26)
27. Pereboom, *Agency, and Meaning in Life, op. cit*., p. 83. [↑](#footnote-ref-27)
28. This may not be nomologically possible, but we can grasp this idea by imagining a backwards light-cone behind each agent, such that the specific facts in that backward light-cone in conjunction with the laws of nature causally determine Plum’s actions but not others. [↑](#footnote-ref-28)
29. Strawsonians are not requiredto grant the Manipulation Premise or the No-Difference Premise, even when existential manipulation is featured in Cases 1–3. They could contend that our practices are not responsive to existential manipulation exemptions or deny that existential manipulation is relevantly like existential determinism. [↑](#footnote-ref-29)
30. For a defence of a similar claim, see Justin Capes, “Mitigating Soft Compatibilism,” *Philosophy and Phenomenological Research*, lxxxvii,3 (2013): 640–663. [↑](#footnote-ref-30)
31. Derk Pereboom, *Living Without Free Will*, *op. cit.*, p. 123. [↑](#footnote-ref-31)
32. Likewise, it may also be possible to imagine a version of this case in which the agents of Population A are determined while the agents of Population B are not. We also think there are a number of ways the responsibility practices of such a world could evolve once the role of determinism is revealed. [↑](#footnote-ref-32)
33. For example, see Patrick Todd, “Strawson, Moral Responsibility, and the ‘Order of Explanation:’ An Intervention,” *Ethics*, cxxvii, 1 (2016): 208–240. [↑](#footnote-ref-33)
34. See, for example, Gary Watson, “Responsibility and the Limits of Evil: Variations on a Strawsonian Theme,” in F. David Schoeman, ed., *Responsibility, Character, and the Emotions: New Essays in Moral Psychology,* (New York: Cambridge University Press). [↑](#footnote-ref-34)
35. Shoemaker takes himself to be the first to offer an extended defence of this view (Shoemaker, “Response-Dependent Responsibility,” *op. cit.*). [↑](#footnote-ref-35)