Erkenntnis
https://doi.org/10.1007/s10670-024-00882-9

ORIGINAL RESEARCH

Compatibilism, Manipulation, and the Hard-Line Reply

Dwayne Moore[1]

Received: 23 April 2024 / Accepted: 9 October 2024

Abstract
Compatibilism is the view that determinism is true, but agents nevertheless possess free will as long as they act from a compatibilist friendly agential structure (i.e., agents want to perform their actions, agents identify with the actions they perform, agents would be responsive to reasons against performing those actions, etc.). The most powerful contemporary objection to compatibilism is the manipulation argument, according to which agents determined to act as they do by the prodding of manipulative neuroscientists are not considered free, so agents determined to act by physical processes operating from the remote past should not be considered free either. One compatibilist response to the manipulation argument is the so-called hard-line reply, according to which compatibilists argue that agents determined by physical processes operating from the remote past are free, so agents determined by manipulative neuroscientists are free as well. In this paper I demonstrate that leading hard-line replies fail, while also introducing a novel argument against the hard-line reply.

1 Introduction

Compatibilism is the view that determinism is true, but agents nevertheless possess free will as long as they act from a compatibilist friendly agential structure (i.e., agents want to perform their actions, agents identify with the actions they perform, agents would be responsive to reasons against performing those actions, etc.). The most powerful contemporary objection to compatibilism is the manipulation argument, according to which agents determined to act as they do by the prodding of manipulative neuroscientists are not considered free, so agents determined to act by physical processes operating from the remote past should not be considered free either. One compatibilist response to the manipulation argument is the so-called hard-line reply, according to which compatibilists argue that agents determined by

✉ Dwayne Moore
dwayne.moore@usask.ca

[1] Philosophy Department, University of Saskatchewan, 9 Campus Drive, Saskatoon, SK S7N 5A5, Canada

Published online: 30 October 2024

physical processes operating from the remote past are free, so agents determined by manipulative neuroscientists are free as well. In this paper I demonstrate that leading hard-line replies fail, while also introducing a novel argument against the hard-line reply.

This paper is divided into five sections. After briefly defining compatibilism, I outline the manipulation argument against compatibilism (§1). I then articulate various replies to the manipulation argument, ultimately emphasizing recent hard-line replies offered by Michael McKenna (2008) and Sofia Jeppsson (2020) (§2). They argue, among other things, that focus on the facts internal to the agential perspective supports the intuition that humans remain free in deterministic and manipulated contexts (§3). I then present three arguments against emphasizing the internal agential perspective in deterministic and manipulated cases (§4). I then provide surplus argumentation against the hard-line reply by introducing the case of the Body Manipulator (§5)—a mad scientist begins by only manipulating Plum's arm into shooting White without Plum desiring to kill White, leaving hard-liners agreeing that Plum is not responsible in this case. The Body Manipulator then manipulates Plum one step back along the causal chain until Plum's brain is manipulated as well, rendering it intuitive that Plum is not responsible when manipulated.

2 Compatibilism and the Manipulation Argument

If determinism is true, then the arrangement of all the conditions of the entire universe at any specific time t (i.e., the arrangement of particles in the universe, the laws of nature, one's brain structure, one's environment, one's background character, personality, upbringing, etc.) causally necessitates all events subsequent to t. Suppose that an event, E, is Ruppert's action of stealing a car at t_1. Then, if determinism is true, the arrangement of all the conditions of the entire universe at a time before t_1 (i.e., Ruppert's brain structure, his being raised in a rough neighborhood, his propensity for stealing, and his finding an unlocked car in an isolated lot) causally necessitates Ruppert's action of stealing the car. Compatibilism, as I use the term here, is the view that determinism is true, yet agents still freely choose how to act, so are properly held morally responsible. On this view, Ruppert, despite being determined to steal the car, nevertheless freely chooses to steal the car, so Ruppert is held morally responsible for his action.

On compatibilism, agents freely choose so long as they act from a so-called "compatibilist-friendly agential structure" or CAS (McKenna, 2008, 142), according to which their agential structures satisfy certain compatibilist conditions. As Derk Pereboom (2001, 101–110) describes it, these conditions include the agent desiring to steal the car (Ayer, 1954), the agent acting from his own character (Hume, 1739), the agent being responsive to countervailing reasons against stealing the car (Fischer & Ravizza, 1998), the agent affirming and identifying with being the type of person who wants to steal the car (Frankfurt, 1971), and the agent being capable of taking both moral and prudential considerations into account (Wallace, 1994). So long as agents act from a CAS, agents act freely and hence are morally responsible, even when their actions are determined. So, though Ruppert is determined to steal the car, Ruppert has reasons and desires leading him to steal the car, he wants to be the type of person who steals the car, he would be responsive to countervailing reasons against stealing the car if they became pertinent, and he is capable of taking his moral considerations into account, so he acts freely and is responsible.

The manipulation argument is perhaps the most trenchant contemporary objection to compatibilism. While there are several versions of the manipulation argument (Stump, 2002, 47–48; Fischer, 1994, 17–20; Kane, 1985, 40; van Inwagen, 1983, 109–110; Mele, 2006, 184–196; Mele, 2019), I shall focus on Derk Pereboom's manipulation argument (Pereboom, 2001, 110–117; Pereboom, 2008; Pereboom, 2014, 76ff). Pereboom imagines a four-stage manipulation argument where in *Case One* a neuroscientist presses a button just prior to Professor Plum's deliberation about killing White, which produces a neural state in Plum's brain that enhances the egoistic reasoning in Professor Plum, which leads Plum to reason in such a way as to decide to kill White.[1] The neuroscientist is subtle, however, so manipulates Plum in a way that ensures that Plum has a CAS—the neuroscientist generates a weak desire in Plum to kill White, so he wants to do it, but is simultaneously capable of being responsive to other reasons, and Plum identifies with his desire to kill White, etc. Pereboom concludes: "Plum's action would seem to satisfy all the compatibilist conditions we examined. But, intuitively, he is not morally responsible because he is determined by the neuroscientists' activities, which are beyond his control" (Pereboom, 2001, 113).

Pereboom then considers *Case Two*, where the neuroscientist creates and programs baby Plum such that Plum will, thirty years later, reason egoistically enough to kill White. Even though Plum possesses a CAS—he wants to kill White, he is responsive to reasons, he identifies with the desire to kill White, etc.—Pereboom concludes that Plum is still not responsible for killing White, since he was still determined to do so by factors beyond his control. In *Case Three* Plum is determined by the rigorous training of his community to tend towards egoistic reasoning. Years later, Plum reasons egoistically enough to want to kill White, though his desires are moderate enough to be reasons responsive and satisfy other compatibilist conditions. Plum kills White, but he is still not responsible because he is still causally determined by factors beyond his control (Pereboom, 2001, 115).

[1] It is worth noting that *Case One* of Pereboom's original four-case manipulation argument leaves open the possibility that the manipulator is "directly producing his every state from moment to moment" (Pereboom, 2001, 111). Some object that this moment-to-moment manipulation undermines Plum's agency in various ways, leaving manipulated Plum relevantly dissimilar to regular Plum, undermining Pereboom's manipulation argument (Fischer, 2004, 156; McKenna, 2008, 149; Demetrio, 2010, 608; Sekatskaya, 2019, 1294). As a result, Pereboom's later version of *Case One* involves a neuroscientist manipulating Plum 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." (Pereboom, 2014, 77). The neuroscientist does not manipulate Plum at every moment of his deliberation, rather the neuroscientist only manipulates Plum before he begins to deliberate about killing Plum. I will follow Pereboom's later *Case One*, where the manipulation occurs via a neuroscientist pressing a button prior to Plum's deliberation, which generates a neural state that yields a desire to kill White, which causes Plum to deliberate about, and ultimately decide to kill White.

Finally, *Case Four* is the normal deterministic case, where Plum is causally determined not by manipulating neuroscientists or cult-like upbringing, but by physical processes operating from the remote past obeying the laws of nature, resulting in typical genetical and cultural factors determining him to reason egoistically enough to follow through on his desire to kill White. In this case too, Pereboom argues, Plum is not responsible, even though he has a CAS. After all, Plum is still completely determined to act by factors beyond his control. In all these cases, Pereboom thinks the responsibility-undermining feature is that Plum is determined to act by factors beyond his control:

> The best explanation for the intuition that Plum is not morally responsible in the first three cases is that his action results from a deterministic causal process that traces back to factors beyond his control. Because Plum is also causally determined in this way in Case 4, we should conclude that here too Plum is not morally responsible for the same reason (Pereboom, 2001, 116).

Since Plum is not responsible when determined by a manipulating neuroscientist, or determined by a cult-like community, Plum is also not responsible when determined by physical processes inevitably running their course from the remote past. It has become common to summarize the manipulation argument in the following three premises:

1. Manipulation Non-Responsibility: Agents determined to act by manipulating beings (neuroscientists, goddesses, etc.) in ways beyond the control of the agent lack moral responsibility for those actions.

2. No Difference: There is no difference relevant to moral responsibility between agents determined to act by manipulating beings in ways beyond their control and agents determined to act by physical deterministic processes in ways beyond their control.

3. Therefore, Determinism Non-Responsibility: Agents determined to act by physical deterministic processes in ways beyond their control lack moral responsibility for those actions.

Plum is not responsible for killing White when determined by a manipulating neuroscientist, whose influence lies beyond Plum's control. And, there is no relevant difference between Plum being determined by a manipulating neuroscientist whose influence lies beyond Plum's control and being determined by physical processes whose influence lies beyond Plum's control. So, Plum is also not responsible for killing White when determined by physical processes whose influence lies beyond Plum's control. If this conclusion is correct, then agents, even if they possess a CAS, are not responsible for their actions when determined by physical processes beyond their control. But compatibilism endorses determinism, according to which agents are determined by physical processes beyond their control, so compatibilism does not deliver moral responsibility. This is a malignant diagnosis, as compatibilists not only desire to preserve responsibility, but compatibilism is the view that agents are free and responsible in determined contexts, so compatibilism would be false if agents actually lack free will and responsibility in determined contexts.

3 Soft-Line and Hard-Line Replies

This manipulation argument against compatibilism stirred up a buzzing beehive of activity, as compatibilists busily set about the task of responding to attacks on their nest. For the most part, compatibilist responses fall into two broad types, which Michael McKenna dubs the hard-line reply and the soft-line reply (McKenna, 2008). The soft-line reply rejects premise (2), claiming there is in fact a relevant difference between manipulation cases and standard determination cases. What is the relevant difference? Some say the difference lies in the presence of the neuroscientist acting as a manipulative agent, while the physical determinism case lacks an agent intentionally compelling Plum (Waller, 2014, 210; Deery & Nahmias, 2017, 1258; Liu, 2022, 547). Others say the distinction lies is in the causal history of the agent, where Plum in *Case One* does not act from a character formed over a long period of time, but Plum in *Case Four* does (Haji & Cuypers, 2007; Mele, 1995, 145–146; Fischer, 2004, 158). Still others insist the distinction lies in the fact that one sole agent is manipulated which sets him apart from how the rest are treated in the manipulation case, whereas everyone is equally determined in the case of physical determinism (Latham & Tierney, 2022). There are other proposals besides.

For every soft-line reply, however, there are readily available rejoinders. With respect to the agent vs. non-agent distinction, critics imagine replacing the manipulator with a non-agential factor that likewise modifies the electrical activity in Plum's brain, such as passing through an electromagnetic field (Mele, 2006, 142) or an unmanned machine (Pereboom, 2001, 115; cp. Mattheson, 2016, 1968ff). With respect to the causal history vs. no causal history distinction, Pereboom offers up *Case Two*, where Plum is manipulated at birth, so has the causal history of a life of character formation as well, but is still not responsible. Al Mele considers the case of Ernie, who is manipulated as a zygote by the goddess Diana to perform a certain action thirty years later. Despite the fact that Ernie has a long causal history, the manipulation precludes his responsibility (Mele, 2006, 188–189). With respect to the sole manipulator vs. the universal determinism distinction, imagine there are billions of manipulators, each tasked with manipulating a different individual. Since everyone is being manipulated now, there is no longer any difference between the quantity of external control occurring in the manipulation and determinism cases. Yet, since everyone is still externally controlled, either by manipulators or determinism, no one seems responsible. To be sure, there are replies to these brief rejoinders, but this suffices to highlight the following important point that advocates of the hard-line reply typically make.

Namely, hard-liners typically grant premise (2), the *No Difference Principle*. That is, they acknowledge that apparent differences between manipulation cases and determinism cases will ultimately be overcome by ever more sophisticated articulations of manipulation cases. As Michael McKenna says: "a crafty incompatibilist can simply alter her manipulation case" (McKenna, 2008, 143), hence taking the soft-line "merely delays the inevitable" (Khoury, 2014, 285). The soft-line reply merely delays resolving the unpalatable binary that either responsibility fails in deterministic cases, or responsibility persists in manipulation cases. Indeed, these crafty alterations just played themselves out. Where the crucial difference was alleged to be agent involvement, incompatibilists imagined manipulation cases *sans* agent involvement. Where the distinction appeared to centre on a lack of historical continuity to the agent, incompatibilists imagined manipulation cases involving a lengthy psychological history to the agent. And so on, and so forth.

Rather than rejecting premise (2), therefore, hard-liners reject premise (1). That is, they say that Plum is morally responsible, even when manipulated by a neuroscientist into killing White. In a sense, they reverse the direction of reasoning. Proponents of the manipulation argument start with the claim that manipulated Plum is not responsible, and then claim there is no relevant difference between manipulated Plum and determined Plum, so determined Plum is not responsible either. Hard-liners start with the claim that determined Plum is responsible, and agree there is no relevant difference between determined Plum and manipulated Plum, so manipulated Plum is responsible as well. Here are some articulations of this hard-line reply:

> A manipulator may succeed, through his interventions, in providing a person not merely with particular feelings and thoughts but with a new character. That person is then morally responsible for the choices and the conduct to which having this character leads. We are inevitably fashioned and sustained, after all, by circumstances over which we have no control. The causes to which we are subject may also change us radically, without thereby bringing it about that we are not morally responsible agents. It is irrelevant whether those causes are operating by virtue of the natural forces that shape our environment or whether they operate through the deliberate manipulative designs of other human agents (Frankfurt, 2002, 27–8).

> Denying premise 1, the compatibilist should welcome the case as one in which the agent is free and responsible. Thus, she should not hide the similarities between the determined and the (properly) manipulated agent. Instead, she should highlight them, arguing that, when the agent is truly manipulated into getting everything that is required by CAS, the intuitive presumption that the agent is not free or morally responsible falters, and with it, the credibility of premise 1 (McKenna, 2008, 144).

> It is, I believe, a straightforward implication of compatibilism that responsible agency can arise from manipulation. The compatibilist holds that there is some deterministic process D that is sufficient for responsible agency; this is simply what the thesis of compatibilism amounts to. There is no reason to think that it is metaphysically impossible that D could be brought about by a manipulation. Given that D is sufficient for responsibility, it will not matter whether D is brought about by natural causes or by a nefarious manipulator or whatever (Khoury, 2014, 285; cp. Fischer, 2011: 269–270; Haas, 2013, 799; Mele, 2021a, 2021b, 303; Tierney, 2013; Liu, 2022, 537; Jeppsson, 2020, 1938; Sekatskaya, 2019, 1292–1293).

Compatibilists think that agents are free and responsible in deterministic scenarios, and hard-liners grant there is no relevant difference between deterministic scenarios and manipulated scenarios, so hard-liners conclude that agents are free and responsible in manipulated scenarios as well.

4 The Agential Perspective

Hard-liners typically provide surplus motivation in support of their position, in hopes of rendering the hard-line position more digestible. Some introduce cases where agents are determined by factors beyond the agent's control, yet we intuitively still hold them responsible (McKenna, 2008, 156–157; McKenna, 2014, 473; Arpaly, 2003, 127–129; Sekatskaya, 2019, 1290). For example, Ann the child watches helplessly as a parent slowly dies of cancer, a traumatic experience which shapes her character as an adult, but we still hold her responsible for her character.

The most common addition to the hard-line reply is emphasis on taking the agential perspective. Recall that compatibilists think that responsibility is established if agents possess a CAS. By maintaining focus on these responsibility-engendering agential conditions, the intuition that agents act freely is augmented: of course, Plum acts freely, after all he wanted to murder White, and he was capable of considering moral reasons to the contrary, and he wanted to be the type of person who murders White. On the contrary, departing from this agential perspective towards the physical, deterministic perspective weakens the intuition that agents are free: of course, Plum does not act freely, after all he is determined to act by factors outside of his control such as a manipulating neuroscientist or physical processing from the remote past. Given that the departure from the agential perspective weakens the intuition that agents are free, and focus upon the agential perspective augments the intuition that agents are free, hard-liners suggest it appropriate to focus on the agential perspective to maintain the intuition that agents are free in these circumstances. Here are several instances of this agential perspective reply:

> A fictional character like the Hologram Doctor, who is presented to us as a fully fleshed-out agent, is perceived differently. As viewers, we are invited to identify with him and the other characters on the show, and to see matters from their point of view. When we do this, our focus naturally moves away from the fact that the Doctor was programmed to do what he does. After all, most of the time, the Doctor faces various decisions and has to decide what to do by weighing his reasons just like the rest of us. When we are invited to identify with him and see things from his point of view, we, too, focus on his reasons, rather than the fact that an engineer once provided him with a program that ultimately determines what he does. When we focus on his reasons for action and which one he ought to choose, we do not get the non-responsibility intuitions that a briefly described thought experiment gives us (Jeppsson, 2020, 1938).

> In this case [4], Plum is just as much like an ordinary moral agent as any one of us. His phenomenology is just as sophisticated. He is capable of feeling

> incredible remorse, or instead ambivalence about killing Ms. White, or alternatively, delicious pride. Furthermore, we can suppose that Plum had a rich history of moral development just like any psychologically healthy person who emerges from childhood into adulthood … The point of attending to these details is to help elicit the intuitions that are friendly to compatibilists—to bring forth the sense that a determined agent of the sort Plum is in Case 4 is a richly complex agent just like any person we might come across (McKenna, 2014, 471; cp. Kapitan, 2000, 97; Double, 1991, 55).

This focus on the agential perspective introduces the contrast between internalist and externalist models of assigning responsibility (Mele, 2021a, 2021b, 249; Mele, 2019, 5ff; Cyr, 2019; Haji & Cuypers, 2007, 359ff; Zimmermann, 1999, 237; Watson, 1999, 360; Frankfurt, 1988, 54). The agential perspective is an internalist model of free will and responsibility, whereby the focus is placed on psychological factors internal to the agent, such as Plum's desire to kill White, Plum's identifying himself as being the type of person who wants to kill White, Plum's responsiveness to any countervailing reasons that may have presented themselves, Plum's capability for feeling guilty, seeking forgiveness and offering justifications, etc. From this internal, agential perspective, Plum looks very "morally articulate" (McKenna, 2008, 151), or is very much like typical free and responsible agents, which bolsters the intuition that he is a free and responsible agent.

Externalists focus on conditions external to the agent, such as the causal history of the agent's character and actions. Imagine Suzie Instant, who is spontaneously created by God several moments ago as a fully-fledged agent, complete with a set of desires and values leading her to kill White. On Internalism, Suzie Instant is responsible for killing White, since she acts from her desires and values. Externalists object: how can she be held responsible for killing White when her values and desires were concocted moments ago by a God, who could have just as easily fashioned a peace-loving Suzie (Haji & Cuypers, 2007; Mele, 2019, 40ff). More relevantly, externalists focus on external, objective factors such as whether an agent is manipulated into, or determined by physical processes into, desiring to kill White. How can Plum be held responsible for killing White when his desires and egoistic reasoning were fashioned by a manipulating neuroscientist, or, determined by physical forces beyond his control?

Hard-liners argue that it is appropriate to focus on the agential perspective rather than the external perspective. Why? There are several reasons. Jeppsson suggests that we should "try to put ourselves in the agent's shoes," rather than making "moral responsibility judgments from a detached causal perspective" (Jeppsson, 2020, 1950).[2] Meanwhile, Frankfurt suggests we should focus on agential conditions relevant to responsibility when assessing freedom and responsibility: "If someone does

[2] Jeppsson, for example, argues that we ought to take the agential perspective when considering relations with friends (Jeppsson, 2020, 1946). There are countervailing intuitions here. Friends often seek an outsider's view, from a psychologist, counselor, or friend, who can provide valuable perspective on what is really going on. Joanne is suddenly being more forthcoming and flirtatious than normal, so her friend provides the external perspective by telling her that the alcohol she has been drinking is affecting her.

something because he wants to do it … it really does not matter how he got that way" (Frankfurt, 2002, 27; cp. Watson, 1999, 360). If these considerations fail, and we should not *prefer* the agential perspective over the external perspective, hard-liners suggest that at the very least we should treat both as *equally* true, resulting in a dialectical stalemate on the issue. As McKenna argues: "an open inventory of both hidden causes and relevant agential properties will end in a dialectical stalemate … this amounts to a victory for the compatibilist, since it is the incompatibilist who is presenting an argument. It is the incompatibilist's burden to make the case stronger than one that closes indecisively" (McKenna, 2008, 148). As McKenna points out, this stalemate is acceptable to hard-liners, as it shows that compatibilists neutralize the manipulation argument attack, leaving compatibilism still standing as a viable model.

5 Problems with the Agential Perspective

Many philosophers reject the hard-line response, arguing that Plum cannot possibly be responsible, given that he is manipulated into acting as he does. As Robert Kane expresses it: "this [the hard-line reply] is a hard line indeed, and one that I think it is also hard to accept" (Kane, 1996, 67). Even compatibilists with hard-line sympathies acknowledge that manipulation arguments reveal "compatibilism's dirty little secret" (Fischer, 2004, 2000 390), and amounts to "taking it on the chin" (McKenna, 2014, 477; cp. Tognazzini, 2014, 358; Haas, 2013, 798). Indeed, it is possible to render stark the unpalatable implications of the hard-line reply by tweaking the thought experiment somewhat. Imagine Jenni, having been raised with some modesty but also desperately wanting to be an actress. She flies to Hollywood to meet with a creepy casting agent who will only give her a job if she sleeps with him, though he wants Jenni to consent to the encounter. In the meeting he hints at what she must do to secure the job, and while she is disturbingly deliberating about it, the casting agent grows impatient at her uncertainty. The casting agent slips a magical pill into her drink, which slightly amplifies the neural activity in the regions of her brain correlated with her desperation for the job. This amplification is enough of an influence to make Jenni focus on her desire for the job, so she chooses to sleep with him, as the casting agent wanted all along. Jenni acts from a CAS at that moment, but it does not seem likely she freely consented to sleeping with him. In this stark case, it is clear that agents possessing a CAS in a manipulated environment are not free or responsible.

In this section I unpack three reasons why the hard-liner's additional emphasis on the agential perspective fails as well. First there is a dual perspective problem. On compatibilism both the agential perspective and the external perspective are true. Hence any suggestion that the agential perspective should be considered the *only* or *most* relevant perspective is belied by the fact that *both* perspectives are equally true, and all facts matter. At this point, compatibilists typically concede that the agential perspective and external perspective are equally true, so compatibilism houses both responsibility-engendering and responsibility-undermining principles within it. However, as seen above, hard-liners think this result implies a dialectical stalemate, which is satisfactory to hard-liners, since it neutralizes the manipulation argument against compatibilism.

However, such a stalemate is not a satisfactory or favourable outcome when considering the wider dialectic. In the wider dialectic compatibilism is in competition with libertarianism (among other models) as the correct model of free will and responsibility. On libertarianism, the responsibility-engendering agential perspective is true, but the responsibility-undermining external perspective is false. After all, libertarians endorse ultimate sourcehood, according to which the source of action lies within agents, and cannot be traced back to causal factors beyond agential control. So, it is not true that agents are determined to act by factors beyond their control, so the responsibility-undermining external perspective is false. Given that libertarianism does not house the responsibility-undermining external perspective, and compatibilism does house the responsibility-undermining external perspective, libertarianism has a dialectical advantage on this point.[3]

There is also a hallucination argument in support of emphasis on the external perspective. To see it, imagine the following modification to the manipulation case:

> *Beach Manipulation*: the mad scientist grows weary of manipulating Plum into performing murderous actions. So, instead of stimulating Plum's brain in a manner in which Plum desires to kill White, the neuroscientist stimulates Plum's brain in a manner that makes Plum think he is standing on a beach on a sunny day. The scientist presses a button, triggering the appropriate neurons in his perceptual systems so that Plum sees a vast ocean before him and feels the sand on his feet. Plum then forms the belief that he is standing on the beach, and decides to start walking on the beach. In reality, however, Plum is sitting on an operating chair in the scientist's lab.

The *Beach Manipulation* case runs parallel to Plum's free will manipulation. In both cases the manipulator presses a button which triggers neural processes in Plum's brain, which then gives rise to an internal subjective experience (i.e., of seeming to be walking on a beach, and of seeming to have freely formed an egoistic desire to kill White). This internal subjective experience then causes Plum to have certain mental states (i.e., deliberations, beliefs) which then causes Plum to act (i.e., by killing White or by starting to walk). In both cases the causal source of the internal subjective experience is incorrectly sensed: it seems to Plum that he freely formed the egoistic desire to kill White while the actual cause was the neuroscientist's brain tampering; and it seems to Beach Plum that an actual beach in the

[3] This is not to say that libertarianism is thereby preferable to compatibilism *tout court*. Libertarianism faces other responsibility-undermining dangers (for example, the luck objection), as compatibilism faces other responsibility-undermining dangers as well (for example, the consequence argument). The point is rather that the responsibility-undermining threat introduced by acceptance of the truth of the external perspective poses problems for compatibilism, but not libertarianism. Hence the issue is not whether there is a stalemate between the manipulation argument and the hard-line reply, rather the issue is that there is not a stalemate between compatibilism and libertarianism on this particular issue. Rather, libertarianism comes out stronger.

external world formed the beach experience while the actual cause was the neuroscientist's brain tampering.

In both cases there is also a stark difference between what the internal perspective presents as true and what the external perspective shows is true. Plum's internal perspective is that his egoistic desire arose naturally, and his deliberations are free in the sense that he can choose what to do, so his action will be caused by factors within his control, but the external perspective shows that his egoistic desire arose by neuroscientific manipulation, which determines that his deliberations will lead to the act of killing White, so his sense that his action is caused by factors within his control is illusory, he is in fact caused to act by factors beyond his control. Beach Plum's internal perspective is that his perception of the beach was caused by an actual beach, so his sense that he is walking on a beach is illusory, he is in fact laying on an operating chair in a lab.

Which perspective should be focused on? Beach Plum's self-deluded, neurochemically concocted hallucination, or the objective reality? Typically, in cases of hallucination, mistaken perception, or deception, it is the objective perspective that trumps the internal perspective. Beach Plum should see that his subjective experience is illusory and false, and that the objective perspective is true. Likewise, if the manipulating neuroscientist stimulates Plum's brain in such a way as to make Plum feel like he freely kills White without being subject to manipulation, this deceived internal perspective should be abandoned in light of the objective fact that Plum is manipulated into killing White.[4]

There is also an ignorance argument in support of favouring the external perspective over the agential perspective. According to the ignorance argument, focusing on the agential perspective is only viable due to our ignorance of the actual hidden causes of our actions. There is no doubt that agents are unaware of the neural causes of their actions—each action is caused by millions of interacting neural processes, each of which possess its own complex biological machinery, and none of which can we consciously introspect. Due to our ignorance of how these neural processes actually determine our behaviour, it appears to agents that they are free and in control of their actions. But this is mere appearance, rooted in ignorance of the actual determining conditions. Pereboom explains that the manipulation argument is expressly intended to make plain the hidden causes actually determining our actions: "Part of the aim of the four-case argument is to foreground in our assessments of moral responsibility the putative hidden causes of our actions, which are not ordinarily

[4] Other similar cases can be imagined. The mad scientist, rather than wanting to make Plum think he freely kills White, wants to make Plum think he had a religious experience. The scientist tweaks Plum's brain such that he feels the presence of a divine being. Plum believes he experienced God, and in that moment commits to devoting his life to this divine being. Should we favour Plum's internal agential reality in this case, or should we favour the objective reality, according to which it was the neuroscientist who stimulating Plum's brain? Seemingly the latter emphasis is appropriate. The general conclusion is again reached: when the internal, agential perspective clashes with the external, objective perspective, the external, objective perspective should be preferred as the most legitimate analysis. Applying this general principle to Plum's killing of White, while from Plum's perspective he reasons freely to the decision to kill White, he is actually being manipulated into doing so, and this responsibility undermining objective reality is the more legitimate analysis.

evident, and in particular the fact that they are deterministic" (Pereboom, 2005, 42). Agents think themselves free because they are not aware of the deterministic processes controlling their behaviour, so the manipulation argument attempts to make plain these hidden causes by putting the influence of a manipulating neuroscientist front and centre.

Consider the following case of ignorant belief. A baby bear sees a huge wolf in the forest, so the cub stands up tall, to scare away the enormous wolf. The wolf is not afraid of the cub, but the wolf cowers away anyway, running back into the forest. The cub feels powerful when she thinks about how she scared off the enormous wolf. Little does the cub know the wolf actually cowered away from fear of the mammoth mother bear standing in behind the cub with dreadful teeth exposed. The cub reached its erroneous conclusion because of her ignorance of the reality hidden from her. Likewise, if an agent reaches the erroneous conclusion that he is free because he is ignorant of how his brain is determining his behaviour, this does not mean the agent is actually free. To see the ignorance argument, consider the case of *Brain Introspecting Plum*:

> *Brain Introspecting Plum*: A mad scientist manipulates Plum's brain such that Plum reasons egoistically enough to form the desire to kill White, and so he kills White. Plum is an unusual specimen, however. He possesses the ability to introspect his own brain states, but cannot introspect his own mental states. So, as soon as the manipulation begins Plum notices the neuroscientist's sudden electrical stimulation of his brain. He observes how this foreign electrical stimulation causes his oxytocin levels to fall, and diminishes the neural activity in his ventromedial prefrontal cortex, which he notices determines a group of neurons to fire, which triggers a musculoskeletal response such that his arm rises and pulls the trigger of a gun, making White fall over and die.

In the *Brain Introspecting Plum* case, the hidden deterministic causes of Plum's behaviour are rendered clear. Revealing this deterministic process increases the intuition that Plum is not responsible for killing White. He has no control over the electrical jolt introduced into his system, nor in how this electrical stimulation shifted his brain chemistry to result in neural processing culminating in musculoskeletal activity in his finger. Since all of this neural processing remains true in the regular case of Plum, it seems like regular Plum should not be considered morally responsible either, given that we are no longer ignorant of all the causal processes leading to White's death.

One can object that *Brain Introspecting Plum* does not accurately reflect compatibilist sentiments. Compatibilists think that Plum has both a deterministic neural cause and a CAS as a cause, but *Brain Introspecting Plum* only has a deterministic neural cause, but cannot apprehend that he also possesses a CAS. It is possible to modify the thought experiment to handle this concern:

> *Mind/Brain Introspecting Plum*: A mad scientist manipulates Plum's brain such that Plum reasons egoistically enough to form the desire to kill White, and so he kills White. Plum is an unusual specimen, however. He possesses the ability to introspect his own brain states, while he can also introspect his own

> mental states. So, as soon as the manipulation begins Plum notices the neuroscientist's sudden electrical stimulation of his brain. He observes how this electrical jolt makes him feel a strengthened sense of selfishness. He then sees how this electrical stimulation causes his oxytocin levels to fall, and diminishes his neural activity in his ventromedial prefrontal cortex, which simultaneously makes him sense his decreasing desire to engage in prosocial behaviour, resulting in a strengthened conviction that he should just kill White. He notices a group of neurons firing while simultaneously resolving to kill White, which triggers a musculoskeletal response such that his arm rises exactly as he wills to raise his arm, and kills White.

In this case Plum can toggle between the CAS of the agential perspective and the deterministic neural perspective, both of which are true on the compatibilist model. But, does Plum seem responsible in this case? It seems not. *Mind/Brain Introspecting Plum* renders clear that Plum's agential processing is actually itself the result of deterministic neural processing in his brain, which he lacks control over. Rather than leading to the mixed result of Plum seeming free because he possesses a CAS though determined, it actually reveals the deeper point that Plum's CAS is itself the product of deterministic neural processes caused by factors beyond his control as well.[5]

Several objections to this conclusion can be gleaned from Jeppsson's writings. Jeppsson imagines the case of Plum being able to introspect his brain states as well, but she imagines different, compatibilist friendly, outcomes to this thought experiment. She suggests Plum may use the knowledge of his manipulation to realize his choice is a sham, so refrain from acting: "we can imagine Plum thinking that whatever he does, his decision will be a sham, so he might just as well do nothing" (Jeppsson, 2020, 1939). She also suggests Plum may worry he lacks any real choice: "he knows that he follows a program, and that given his program and the environment in which he finds himself only one action is possible" (Jeppsson, 2020, 1940). She then suggests Plum may use the manipulation to excuse his blame: "the fact that he is manipulated thus supplies him with a reason to commit the murder and take the money" (Jeppsson, 2020, 1940). Despite all these new considerations, Jeppsson

[5] One can object that this argument is too strong, as it threatens not only deterministic views of free will, but indeterministic views of free will as well. In fact, it may threaten all physicalist models of free will, as physicalists grant that agential states supervene upon, hence are determined by, brain processes over which agents lacks control. In response, I agree this jeopardizes any model of free will that claims that agential states are completely brought about by factors over which agents have no control. This is, after all, the central insight of the manipulation argument. This includes compatibilist models that claim that agential states are completely brought about by causes in the remote past over which the agent lacks control, and I have argued elsewhere that this includes any compatibilist (Moore, 2023) any libertarian models (Moore, 2021) that claim that agential states are completely brought about by brain processes over which agents lacks control. What options remain? I argue that nonreductive physicalists may be able to deliver libertarian free will in a manner that overcomes this objection (Moore, *Forthcoming*). There are also a variety of agent causal models of libertarian free will that may overcome this objection. Although I argue that compatibilists cannot overcome this objection, it is possible to imagine nonreductive physicalist models of compatibilism that may succeed. I would like to thank an anonymous reviewer for pressing this point.

concludes that Plum must still focus on his agential perspective (i.e., his reasons for and against killing White, and the decision he will make), lessening the focus on his manipulation (Jeppsson, 2020,1939–1940). For Jeppsson, Plum sees the decision he must make, and his reasons for and against the murder are at the foreground, so much so that the side reality of noticing his manipulation is relegated to being a less important issue.

There are several problems with Jeppsson's interpretation. First, if it is true that Plum's knowledge of his determined condition leads him to realize his choice is a sham, then this serves as an objection to compatibilism. After all, compatibilists think everyone is determined, so the concession that Plum realizing his determined state leads to choice being a sham actually supports the incompatibilist intuition that determinism is not compatible with free choice. Compatibilists also typically reject the view that fatalism follows from determinism, so the suggestion that Plum may just do nothing once he realizes he is following a determined program once again supports the incompatibilist intuition that determinism undermines free will. Likewise Plum absolving himself of responsibility due to his determined state also supports the incompatibilist intuition that determinism undermines free will and responsibility. Compatibilists think agents remain responsible, even after knowing of their determined state, so a compatibilist should think *Mind/Brain Introspecting Plum* does so as well. It is also not appropriate to focus only on the agential perspective, given that both the agential perspective and the external perspective are true. Since both perspectives are true, Plum should focus on both perspectives. Indeed, if Plum should just focus on one of two true perspectives, perhaps he should just focus on the external perspective. This leads to the result that, despite Plum occasionally sensing his reasons and desires, he focuses on how the neuroscientist's manipulation determines his brain processes to respond, leading him to conclude he is not free.

6 The Body Manipulator

In the last section I responded to the agential perspective argument, a leading motivation supporting the hard-line reply. In this section I construct a novel argument against the hard-line reply, called the Body Manipulation argument. It begins by noting that Pereboom begins his manipulation argument with an outlandish case of direct manipulation in *Case One*, where it is supposed to be intuitively obvious that Plum is not free when directly manipulated. Surprisingly, the hard-liner bites the bullet, and accepts that Plum is free in this outlandish case. Thus, in this section I expand the manipulation argument to begin with new, even more outlandish cases of manipulation; cases where even the staunchest hard-liner will not claim Plum is responsible. Then, as in the original manipulation argument, I work my way back from the outlandish cases to Pereboom's *Case One*, which then leads to the typical case of determinism in *Case Four*. Consider this case:

> *Arm Manipulation*: A mad scientist wants White to die, and the scientist wants to take credit for her crime. But it is beneath a scientist of her intellect to simply shoot White as a normal hoodlum would, so she concocts an elaborate

> plan. She fastens a technological device to Plum's arm, such that when she presses a button, a series of electrical signals cause Plum's arm muscles to contract in such a way as to make Plum's arm rise and pull the trigger of a gun, killing White.

Even though Plum is part of the *Arm Manipulation* case, Plum does not freely kill White. Rather, Plum's hand is being used as a tool for the mad scientist to commit the murder. Even the hard-liner is on board here: Plum did not act from a CAS, Plum did not even act, and Plum did not even have a CAS—he has no desire to kill White, nor does he intend to kill White, nor did he perform the act of pulling the trigger, so Plum is not responsible. Rather, Plum stands there bewildered as his arm is shocked into involuntary spasms. It is the mad scientist who is responsible for killing White. Consider then the following case:

> *Motor Neuron Manipulation*: A mad scientist wants White to die, and the scientist wants to take credit for her crime. But it is beneath a scientist of her intellect to simply shoot White as a normal hoodlum would, so she concocts an elaborate plan. She fastens a technological device in Plum's upper motor neurons, such that when she presses a button, a series of electrical signals cause Plum's upper motor neurons to fire, causing appropriate lower motor neurons to fire, causing Plum's arm muscles to contract in such a way as to make Plum's arm rise and pull the trigger of a gun, killing White.

Here the mad scientist is simply doing what she already did, namely, stimulating a part of Plum's body such that it moves as she wishes, though she starts the manipulative intervention upstream a bit in Plum's motor cortex. The central intuition still holds: Plum is not responsible for killing White. His body is still being used as a tool for the mad scientist to make a gun kill White. Hard-liners agree, as Plum still lacks a CAS, since the manipulation occurs in a brain region causally posterior to the brain regions responsible for deliberation and CAS formation. It is still the mad scientist who is to blame. Let us consider this case then:

> *Cortex Manipulation*: A mad scientist wants White to die, but she is tired of taking the blame, so she wants Plum to be blamed this time. So, rather than fastening a technological device in his upper motor neurons, such that Plum will not form a CAS, she fastens a technological device further back in Plum's cerebral cortex. When she presses a button, a series of electrical signals cause Plum's oxytocin levels to fall, and diminishes neural activity in his ventromedial prefrontal cortex. This intervention not only causes Plum to reason egoistically and form the desire the murder White, but simultaneously causes his upper motor neurons to fire, causing appropriate lower motor neurons to fire, causing Plum's arm muscles to contract in such a way as to make Plum's arm rise and pull the trigger of a gun, killing White.

Again, the mad scientist is simply doing what she already did—stimulating a part of Plum's body such that it moves as she wishes—though she starts the manipulation earlier in Plum's brain. If Plum was not responsible when the scientist triggered Plum's body to move before, Plum should not be responsible now either, since the scientist is simply doing the same thing, but in a different region of Plum's body this time. Indeed, Plum seems even less responsible now, as he is now doubly manipulated. Not only does the mad scientist manipulate Plum's body into killing White, but she also stimulates Plum's brain in such a way as to make Plum want to commit the murder. The mad scientist is doubly responsible—not only did she use Plum's body to kill White, but she also manipulated his mind into having the psychological profile required for the police to blame Plum instead of her. But the *Cortex Manipulation* case is equivalent to Pereboom's *Case One*, where the mad scientist also manipulates Plum's brain into forming a CAS and killing White. So, if Plum is not responsible in *Cortex Manipulation*, then Plum is not responsible in Pereboom's *Case One*, undermining the hard-liner's claim that Plum is responsible in *Case One*.

The hard-liner may object, however. Plum has a CAS in *Cortex Manipulation*, so Plum now satisfies the conditions required for responsibility, and Plum is to blame for killing White. Indeed, this is exactly what the mad scientist is relying on. The response to this concern is to move more slowly from the *Motor Neuron* case (where Plum is clearly not responsible) back to the *Cortex Manipulation* case, where this slower movement preserves the former intuition that Plum is not responsible through to the final case. Imagine the mad scientist starting from the *Motor Neuron Manipulation* case, but then moving her manipulation backwards along the antecedent causal chain one neuron at a time. Here she simply stimulates the neuron one step prior to the ones she had previously stimulated. Plum once again follows through on killing White, but is he responsible? No. If Plum is not responsible when the manipulation starts with stimulating these motor neurons, surely Plum is not responsible when stimulating one neuron prior in his neural wiring. And, if he is not responsible when the scientist stimulates one neural firing earlier, then why would Plum be responsible if the manipulation starts two neurons earlier? And so on. The intuition is the same at each moment: if Plum kills White due to the neuroscientist electrochemically stimulating a region of Plum's body, which then deterministically ensures that Plum kills White, then Plum is not responsible, no matter which region of Plum's body or brain the neuroscientist begins her prodding in, and no matter whether this prodding also has the effect of generating a CAS or not.

A similar conclusion is reached when considering the agential structure ensuing from the neuroscientist's backwards-step-wise manipulation. Imagine the neuroscientist manipulates Plum's brain at the terminal point in Plum's CAS. Plum does not desire to kill White, nor does he deliberate about whether to kill White, but due to the manipulation he forms a sudden intention out of nowhere to kill White, which causes him to kill White. White is surprised at how hastily he chooses to pull the trigger. Is Plum responsible now? For incompatibilists the answer is no—his brain and agential structure, hence ensuing behaviours, are determined by factors beyond his control, so he is not responsible. Hard-liners likely answer no as well. If anything, Plum only has a partial-CAS. He did form the intention to kill White, which caused him to murder White, which is incriminating. But he lacks key features of a CAS: he did not deliberate in a reasons responsive way, he did not form a desire to kill White, he did not identify with his desire to kill White, etc. Indeed, since his reasons did not cause the killing behaviour, the killing would not be considered Plum's action (according to the Causal

Theory of Action), so Plum is not responsible. Plum's sudden intention to shoot White rises up in him like an uncontrolled reflex reaction, like how being startled would.

Now imagine that the manipulator begins her electrochemical intervention one step earlier in Plum's brain, and hence in Plum's agential structure. Perhaps the neuroscientist enhances his egoistic reasoning, and suppresses his moral reasoning to such an extent as to render him non-reasons responsive. Now Plum notices a desire to murder White, and finds himself noticing reasons supportive of the murder popping into his mind. No countervailing reasons arise, so he forms the intention to kill White, and he does. Is Plum responsible now? For incompatibilists the answer remains obvious—his brain and hence agential states and ensuing behaviours are still determined by the prodding neuroscientist, it does not matter which set of neurons she starts her intervention on, Plum is not responsible. Hard-line compatibilists may begin to splinter off into differing views at this point. Plum did desire to kill White, and formed egoistic reasons for doing so, which led to the intention and ensuing act of killing White. Plum clearly acts now, as the murder is the result of his reasons and desires. On the other hand, he still only possesses a semi-CAS. He did not identify with his reasons for murdering White, and his moral reasoning was suppressed such that he was unable to deliberate in a reasons responsive manner. So, Plum still lacks several essential features of a CAS, so is not responsible.

Now imagine that the manipulator begins her electrochemical intervention in a slightly different region of Plum's brain, which modifies Plum's agential structure differently. Now the neuroscientist enhances his egoistic reasoning, but does not suppress his moral reasoning. Now Plum senses a desire to murder White, and finds himself with motivating reasons in support of killing White. A host of countervailing moral reasons enter his deliberations, but they are not strong enough to overcome his carefully augmented egoistic reasoning, so he forms the intention to kill White, and does so. Is Plum responsible now? For the incompatibilist, the answer remains simple: Plum's behaviour and agential structure is still determined by factors beyond his control, so he is still not responsible, it does not matter where the neuroscientist begins tweaking Plum's brain. Hard-liners, however, remain divided. He possesses an almost-CAS: he has the desire to kill White and the egoistic reasoning supporting it, and he is capable of moral reasoning and reasons responsiveness, which are many of the elements in a CAS. He does not, however, identify with his desire to kill White, so lacks a crucial ingredient of the CAS, so he may not be responsible.

Now imagine one final case: the neuroscientist begins her electrochemical intervention in the same region as she did in the almost-CAS, and Plum begins to desire to kill White and reason egoistically enough to justify it. However, the neuroscientist then locates the region of Plum's brain responsible for him identifying with this desire to kill White, and she stimulates that region until Plum desires to be the type of person who desires to kill White, so Plum forms the intention to kill White, and kills White. Is Plum responsible now? Incompatibilists still say no, he is still determined by factors beyond his control, it does not matter which region of his brain the manipulation begins in. Many hard-line compatibilists will say yes at this point. Plum has a full-CAS now: he desires to kill White, reasons egoistically enough to justify the desire, is responsive to countervailing reasons, and wants to be the type of person who kills White, so he is responsible.

But this result is counterintuitive. There is only one small difference between this case and the almost-CAS case where many hard-liners said Plum is not responsible. The only difference is that the neuroscientist completes an additional manipulation, stimulating Plum's brain such that Plum wants to be the type of person to kill White. Is this sufficient to take Plum from non-responsibility to responsibility? Incompatibilists do not think so. Why would an additional electrochemical intervention in Plum's brain suddenly make Plum responsible? Some hard-liners will agree here as well, presumably those who do not think Frankfurt's compatibilist conditions are the sufficient condition on responsibility. Those hard-liners perhaps thought Plum became responsible during an earlier intervention, perhaps at the semi-CAS stage, or the almost-CAS stage. Even hard-liners, then, will not definitively agree that Plum is responsible now.

This full-CAS case is similar to Pereboom's *Case One*, where Plum is also manipulated by a mad scientist such that Plum has a CAS, and Plum murders White. So, if hard-liners are now divided, or undecided, about this full-CAS case, hard-liners should also have weakened resolve to accept that Plum is responsible in Pereboom's *Case One*. In addition, this argument reveals the larger intuition against the hard-line view: once it is clear that Plum's behaviour, as well as his agential structure, including whether he has a CAS or not, or has some sort of partial-CAS (or, perhaps, an intermittent CAS, where the CAS flickers on and off during the deliberation) is completely controlled by the locale and manner in which the neuroscientist begins the manipulation in Plum's body or brain, we increasingly see that Plum is like a puppet on a string, and is not responsible for what the manipulator makes him do.

An objection to this line of reasoning comes from Maria Sekatskaya's writings. Sekatskaya (2019) imagines expansions to Pereboom's Four Case argument as well, though she begins in the opposite direction, with cases where the agent seems responsible, before working her way back to *Case One*. She begins with a responsibility-friendly Case A, where a colleague provides new information to Plum, which makes Plum desire to kill White, and he does kill White (Sekatskaya, 2019, 1290). In Case B a neuroscientist pretends to be the new colleague, and gives Plum the same new information, which makes Plum desire to kill White, and he does so. Plum is still considered responsible in Case B as well. Next the neuroscientist manipulates Plum's perceptual system such that Plum hallucinates that a colleague provides new information, where this hallucination makes Plum desire to kill White. Plum is still responsible for his action here. Next the neuroscientist implants false memories, making Plum falsely recall the colleague providing new information, which makes Plum desire to kill White. Plum is still responsible here. Finally, in Case E, the neuroscientist skips the theatrics of false memories and false perceptions, and just gives rise to Plum's desire to kill White. Plum is still responsible here (Sekatskaya, 2019, 1292), and this case is similar to Pereboom's *Case One*, showing that Plum is still responsible in *Case One*, which undermines the manipulation argument.

In response, at worst the result is a stalemate between the cases. Sekatskaya's cases begin with the appearance of responsibility, then she tries to preserve this responsibility back to Pereboom's *Case One*. I begin here with the appearance of non-responsibility, then preserve this non-responsibility intuition back to Pereboom's *Case One*. The final result need not end in this stalemate of competing cases, however. Rather, the cases I list here are intended to highlight the important litmus test for determining responsibility, and this litmus test can be applied to Sekatskaya's cases as well. The litmus test for determining responsibility remains the same: if forces beyond the agent's control completely cause the agent to act, then the agent is not responsible. The cases presented here are meant to draw this intuition out. Of course, if a mad scientist completely causes Plum's limbs to move, Plum is not responsible. Indeed, if a mad scientist stimulates Plum in a million different ways (i.e., causing his leg to kick, causing his eyes to cry, causing his heart to race, etc.), Plum is not responsible for any of it because his body is completely caused to behave by forces beyond his control, namely, the whims of the mad scientist. If the mad scientist decides to tamper with Plum's brain rather than the other million parts of Plum's body that she has already tampered with, it is still the case that Plum is not responsible because he is completely caused to act by forces beyond his control, even if the mad scientist's manipulation also gives rise to a CAS in Plum.

This litmus test can be applied to Sekatskaya's cases as well. Sekatskaya wonders if Plum is responsible in Case A, where a colleague influences Plum to kill White. The answer is gleaned from sorting out this simple question: was Plum completely caused to act by forces beyond his control in Case A? If so, then he is not responsible. If not, then he may be responsible. Unfortunately, Case A is under-described in this regard, as it does not clearly state whether Plum is completely caused to act by forces beyond his control or not. If we add this litmus test to Case A, however, our intuitions become clearer. Imagine Case A_1, where a colleague discusses White with Plum such that Plum forms a desire to kill White and does so, and as it turns out Plum's decision to kill White was completely caused by factors beyond Plum's control (perhaps by the colleague having a device that scans Plum's brain patterns as the discussion progresses, and alerts the colleague when Plum's brain indicates Plum's strong desire to kill White, and notices that Plum's brain patterns are completely deterministic). The intuition in Case A_1 is that Plum is not responsible, given that he is completely caused to act by factors beyond his control, leaving it true that he did not make up his own mind on the matter.

Now consider Case A_2, where a colleague discusses White with Plum such that Plum forms a desire to kill White and does so, and in this case Plum's decision to kill White was not completely caused to act by factors beyond Plum's control (perhaps by the colleague having a device that scans Plum's brain patterns as the discussion progresses, and alerts the colleague when Plum is perfectly undecided about how he feels about Plum, and his brain processing is indeterministic such that different outcomes are possible, leaving the decision totally up to Plum). The intuition in Case A_2 is that Plum is responsible, given that he is not completely caused to act by factors beyond his control, leaving him able to make up his own mind on the matter. In both cases the key question is whether Plum is completely caused to act by forces beyond his control or not. In Pereboom's manipulation cases, as with determinism, the answer is yes, so it becomes clear that Plum is not responsible once this is clarified.

In summary, the hard-line response to the manipulation argument is unsuccessful. The hard-liner's focus on the agential, responsibility-engendering perspective fails because the objective, responsibility-undermining perspective is also true on compatibilism, and there is good reason to emphasize this objective, responsibility-undermining perspective. Moreover, introducing new manipulation cases, such as body manipulation cases, then slowly working back to regular cases of manipulation, highlights the fact that manipulated agents lack control over both their behaviour and whether they possess a CAS or not, strengthening the intuition that manipulated agents are not responsible. Given the failure of the hard-line response, it remains open for compatibilists to retreat back towards the soft-line response, but that response has its share of difficulties as well. Compatibilists, then, would be well served to consider other models of free will and responsibility.

Declarations

Conflict of interest No conflicts of interest.

References

Arpaly, N. (2003). *Unprincipled virtue: An inquiry into moral agency*. Oxford University Press.

Ayer, A. J. (1954). *Philosophical essays*. Greenwood Press.

Cyr, T. (2019). Why compatibilists must be internalists. *The Journal of Ethics, 23*(4), 473–484.

Deery, O., & Nahmias, E. (2017). Defeating manipulation arguments: interventionist causation and compatibilist sourcehood. *Philosophical Studies, 174*(5), 1255–1276.

Double, R. (1991). *The non-reality of free will*. Oxford University Press.

Fischer, J. (1994). *The metaphysics of free will: an essay on control*. Blackwell.

Fischer, J., & Ravizza, M. (1998). *Responsibility and control: A theory of moral responsibility*. Cambridge: Cambridge University Press.

Fischer, J. (2000). Responsibility, History and Manipulation, *Journal of Ethics, 4*, 385–391.

Fischer, J. (2004). Responsibility and manipulation. *The Journal of Ethics, 8*(2), 145–177.

Fischer, J. (2011). The zygote argument re-mixed. *Analysis, 71*(2), 267–272.

Frankfurt, H. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy, 68*(1), 5–20.

Frankfurth, H. (1988). *The importance of what we care about*. Cambridge University Press.

Frankfurt, H. (2002). Reply to John Martin Fischer. In Sarah Buss & Lee Overton (Eds.), *The contours of agency: Essays on themes from Harry Frankfurt* (pp. 27–32). The MIT Press. https://doi.org/10.7551/mitpress/2143.003.0020

Haas, D. (2013). In defense of hard-line replies to the multiple-case manipulation argument. *Philosophical Studies, 163*(3), 797–811.

Haji, I., & Cuypers, S. (2007). Magical agents, global induction, and the internalism/externalism debate. *Australasian Journal of Philosophy, 85*(3), 343–371.

Hume, D. (1978). *A treatise of human nature*. Oxford: Oxford University Press.

Jeppsson, S. (2020). The agential perspective: A hard-line reply to the four-case manipulation argument. *Philosophical Studies, 177*(7), 1935–1951.

Kane, R. (1985). *Free will and values*. Albany: State University of New York Press.

Kane, R. (1996). *The significance of free will*. Oxford University Press.

Kapitan, T. (2000). Autonomy and manipulated freedom. *Philosophical Perspectives, 14*(s14), 81–104.

Khoury, A. (2014). Manipulation and mitigation. *Philosophical Studies, 168*(1), 283–294.

Latham, A., & Tierney, H. (2022). Defusing existential and universal threats to compatibilism: A strawsonian dilemma for manipulation arguments. *Journal of Philosophy, 119*(3), 144–161.

Liu, X. (2022). Manipulation and machine induction. *Mind, 131*(522), 535–548.

McKenna, M. (2008). A hard-line reply to pereboom's four-case manipulation argument. *Philosophy and Phenomenological Research, 77*(1), 142–159.
McKenna, M. (2014). Resisting the manipulation argument. *Philosophy and Phenomenological Research, 89*(2), 467–484.
Mele, A. (1995). *Autonomous agents: From self-control to autonomy*. Oxford University Press.
Mele, A. (2006). *Free will and luck*. Oxford University Press.
Mele, A. (2019). *Manipulated agents: A window to moral responsibility*. Oxford University Press.
Mele, A. (2021a). Manipulated agents: Precis. *Criminal Law and Philosophy, 15*, 249–253.
Mele, A. (2021b). Manipulated agents: Replies to critics. *Criminal Law and Philosophy, 15*, 299–309.
Moore, D. (2021). Libertarian free will and the physical indeterminist luck objection. *Philosophia, 50*(1), 159–182.
Moore, D. (2023). Lemos on the physical indeterminism luck objection. *Philosophia, 51*, 1459–1477.
Moore, D. (forthcoming). A Nonreductive Physicalist Libertarian Free Will. *Inquiry*. https://doi.org/10.1080/0020174X.2023.2166982
Pereboom, D. (2001). *Living without free will*. Cambridge University Press.
Pereboom, D. (2005). Defending hard incompatibilism. *Midwest Studies in Philosophy, 29*, 228–247.
Pereboom, D. (2008). A hard-line reply to the multiple-case manipulation argument. *Philosophy and Phenomenological Research, 77*(1), 160–170.
Pereboom, D. (2014). *Free will, agency, and meaning in life*. Oxford: Oxford University Press.
Sekatskaya, M. (2019). Double defence against multiple case manipulation arguments. *Philosophia, 47*(4), 1283–1295.
Stump, E. (2002). Control and causal determinism. In Sarah Buss & Lee Overton (Eds.), *The contours of agency: essays on themes from harry frankfurt* CAPITAL H AND CAPITAL F IN HARRY FRANKFURT (pp. 33–60). The MIT Press. https://doi.org/10.7551/mitpress/2143.003.0005
Tierney, H. (2013). A maneuver around the modified manipulation argument. *Philosophical Studies, 165*(3), 753–763.
Tognazzini, N. A. (2014). The structure of a manipulation argument. *Ethics, 124*(2), 358–369.
Van Inwagen, P. (1983). *An essay on free will*. Clarendon Press.
Wallace, R. (1994). *Responsibility and the moral sentiments*. Harvard University Press.
Waller, R. (2014). The threat of effective intentions to moral responsibility in the zygote argument. *Philosophia, 42*(1), 209–222.
Watson, G. (1999). Soft libertarianism and hard compatibilism. *The Journal of Ethics, 3*(4), 351–365.
Zimmerman, D. (1999). Born yesterday: personal autonomy for agents without a past. *Midwest Studies in Philosophy, 23*(1), 236–266.