



# Why Compatibilists Must Be Internalists

Taylor W. Cyr<sup>1</sup>

Received: 4 June 2019 / Accepted: 4 September 2019 / Published online: 10 September 2019  
© Springer Nature B.V. 2019

## Abstract

Some compatibilists are *internalists* (or *structuralists*). On their view, whether an agent is morally responsible for an action depends only on her psychological structure at that time (and not, say, on how she came to have that structure). Other compatibilists are *externalists* (or *historicists*). On their view, an agent's history (how she came to be a certain way) can make a difference as to whether or not she is morally responsible. In response to worries about manipulation, some internalists have claimed that compatibilism requires internalism. Recently, Alfred Mele has argued that this internalist response is untenable. The aim of this paper is to vindicate the claim that compatibilism requires internalism, showing where Mele's argument goes wrong along the way.

**Keywords** Compatibilism · Externalism · Internalism · Alfred Mele · Moral responsibility

## 1 Introduction

While compatibilists about moral responsibility agree with one another that determined agents may nevertheless be morally responsible, they disagree about whether (and how) an agent's history affects her moral responsibility.<sup>1</sup> Compatibilists may be roughly divided into two camps. In the first camp are those that may be called *internalists* (or *structuralists*). On their view, whether an agent is morally responsible for an action depends only on her psychological structure at that time (and not, say, on how she came to have that structure). *Externalists* (or *historicists*), by contrast, build

<sup>1</sup> In this paper, I am concerned with direct (or non-derivative) moral responsibility. Since everyone should agree that there is a distinction between direct and indirect moral responsibility, and since this distinction is a historical one, everyone should think that, in cases of *indirect* moral responsibility, whether or not an agent is morally responsible can depend on non-structural features of an agent. For more on this point, see McKenna (2012: 156).

✉ Taylor W. Cyr  
taylor.w.cyr@gmail.com

<sup>1</sup> Samford University, 800 Lakeshore Drive, Birmingham, AL 35229, USA

historical conditions into their accounts of moral responsibility, such that, on their accounts, an agent's history (how she came to be a certain way) can make a difference as to whether or not she is morally responsible.<sup>2</sup>

Arguably the biggest challenge for internalists is a challenge based on cases of manipulation. The worry is that, so long as one's account of moral responsibility is a time-slice (internalist) notion, it is possible for one's account's allegedly sufficient conditions on moral responsibility to be satisfied by agents who have been manipulated—and who, thus, appear not to be morally responsible for their behavior. One popular response by internalists, including such influential compatibilists as Harry Frankfurt and Gary Watson, is to claim that a commitment to compatibilism requires a commitment to internalism. Recently, Mele (2016, 2019, chapter 4) has argued that this internalist response is untenable. The aim of this paper is to vindicate the claim that compatibilism requires internalism, showing where Mele's argument goes wrong along the way.

I will proceed as follows. In Sect. 2, I will reproduce some important, representative, and widely cited passages by internalists about the connection between compatibilism and internalism. Then, in Sect. 3, I will summarize Mele's challenge to the idea that compatibilism requires internalism. In Sect. 4, I lay some important groundwork for my reply to Mele by invoking a relevant distinction that Mele overlooks in this context. Finally, in Sect. 5, I reply to Mele with an argument for the conclusion that compatibilists must be internalists, highlighting its implications for certain cases of manipulation.

## 2 Some Internalists: Frankfurt and Watson

Frankfurt's internalist account has been the most influential of its type, and it is worth starting with an early statement of Frankfurt's view on agents' histories:

To the extent that a person identifies himself with the springs of his actions, he takes responsibility for those actions and acquires moral responsibility for them; moreover, the questions of how the actions and his identifications with their springs are caused are irrelevant to the questions of whether he performs the actions freely or is morally responsible for performing them (Frankfurt 1988: 54).

If how an agent's action or her identification with that action is caused is irrelevant to whether the agent is morally responsible, then it would not matter (concerning the agent's moral responsibility) whether an agent had been manipulated by another agent into having that identification and performing that action. But many

---

<sup>2</sup> It is worth noting that this disagreement is not an in-house debate among compatibilists (though it is often discussed as though it were): even *incompatibilist* accounts of moral responsibility (i.e., libertarian accounts) may be divided into internalist and externalist camps. That said, I will limit my focus to compatibilist accounts here.

philosophers find this counterintuitive, and some have pressed Frankfurt on this. In response, Frankfurt says:

If someone does something because he wants to do it, and if he has no reservations about that desire but is wholeheartedly behind it, then—so far as his moral responsibility for doing it is concerned—it really does not matter how he got that way. One further requirement must be added...: the person's desires and attitudes have to be relatively well integrated into his general psychic condition. Otherwise they are not genuinely his...As long as their interrelations imply that they are unequivocally attributable to him...it makes no difference—so far as evaluating his moral responsibility is concerned—how he came to have them (Frankfurt 2002: 27).

And specifically concerning cases of manipulation, Frankfurt continues:

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 deliberately manipulative designs of other human agents (Frankfurt 2002: 28).

As is clear from this last quotation, Frankfurt's view is that there is no relevant difference between being caused to meet his internalist conditions on moral responsibility by another's agency, on the one hand, or by "natural forces," on the other. And if he is right that there is no relevant difference, then internalism seems to follow from compatibilism, since compatibilists agree that agents determined by the distant past and the laws of nature ("natural forces") may nevertheless be morally responsible.

Now consider the following passage from Gary Watson, another widely discussed and influential internalist:

For the compatibilist, the constitutive conditions of free agency do not conceptually depend on their origins. In this sense, free and responsible agency is not an historical notion. Consequently, compatibilism is committed to the conceptual possibility that free and responsible agents, and free and responsible exercises of their agency, are products of super-powerful designers. For consider any compatibilist account of the conditions of free agency, C. It is possible for C to obtain in a causally deterministic world. If that is possible, then it is possible that a super-powerful being intentionally creates a C-world, by bringing about the relevant antecedent conditions in accordance with the relevant laws. This possibility follows from the general point that the conditions of responsibility do not necessarily depend upon their causal origins (Watson 1999: 360–361).

In this passage, Watson is making a similar point to the one Frankfurt went on to make in the passages considered above. Roughly, if compatibilism is true, then it must be possible for agents to satisfy any compatibilist account of the conditions of moral responsibility (and free agency) despite being set up by a super-powerful being.<sup>3</sup>

### 3 Mele's Criticisms

To understand Mele's criticisms of these lines of thought, it will be helpful to consider one of Mele's cases of manipulation, which is worth quoting at some length. The passage begins with descriptions of two agents, Beth and Chuck, and then tells how Beth is manipulated (taken from Mele 2016: 72–74):

Beth is one of the kindest, gentlest people on Earth. She was not always that way, however. When she was a teenager, she came to view herself, with some justification, as self-centered, petty, and somewhat cruel. She worked hard to improve her character, and she succeeded. Beth is an extremely kind and generous person who for many years has devoted a great deal of time and energy to helping needy people in her community and the local Girl Scouts....

Chuck enjoys killing people, and he "is wholeheartedly behind" his murderous desires, which are "well integrated into his general psychic condition." (Frankfurt 2002: 27) When he kills, he does so "because he wants to do it" (Frankfurt 2002: 27), and "he identifies himself with the springs of his action." (Frankfurt 1988: 54) When he was much younger, Chuck enjoyed torturing animals, but he was not wholeheartedly behind this. These activities sometimes caused him to feel guilty, he experienced bouts of squeamishness, and he occasionally considered abandoning animal torture. However, Chuck valued being the sort of person who does as he pleases and who unambivalently rejects conventional morality as a system designed for and by weaklings. He freely set out to ensure that he would be wholeheartedly behind his torturing of animals and related activities, including his merciless bullying of vulnerable people, and he was morally responsible for so doing. One strand of his strategy was to perform cruel actions with increased frequency in order to harden himself against feelings of guilt and squeamishness and eventually to extinguish the source of those feelings. Chuck strove to ensure that his psyche left no room for mercy. His strategy worked (Mele 1995: 162–163, 2006: 171).

When Beth...crawled into bed last night she was an exceptionally sweet person, as she always had been. Beth's character was such that intentionally doing anyone serious bodily harm definitely was not an option for her: her char-

<sup>3</sup> One difference worth noting is that Frankfurt seems to have cases of mid-life manipulation in mind (what Mele calls "radical reversals"), whereas Watson is imagining that the powerful being sets the world in motion (which would be a case of what Mele calls "original design"). I agree with Mele that there is an important asymmetry between such cases, and I'll return to this point in Sect. 5.

acter—or collection of values—left no place for a desire to do such a thing to take root. Moreover, she was morally responsible, at least to a significant extent, for having the character she had. But Beth awakes with a desire to stalk and kill a neighbor, George. Although she had always found George unpleasant, she is very surprised by this desire. What happened is that, while Beth slept, a team of psychologists that had discovered the system of values that make Chuck tick implanted those values in Beth after erasing hers. They did this while leaving her memory intact, which helps account for her surprise. Beth reflects on her new desire. Among other things, she judges, rightly, that it is utterly in line with her system of values. She also judges that she finally sees the light about morality—that it is a system designed for and by weaklings. Upon reflection, Beth “has no reservations about” her desire to kill George and “is wholeheartedly behind it.” (Frankfurt 2002: 27) Furthermore, the desire is “well integrated into [her] general psychic condition.” (Frankfurt 2002: 27) Seeing absolutely no reason not to stalk and kill George, provided that she can get away with it, Beth devises a plan for killing him, and she executes it—and him—that afternoon. That she sees no reason not to do this is utterly predictable, given the content of the values that ultimately ground her reflection. Beth “identifies [herself] with the springs of her action” (Frankfurt 1988: 54), and she kills George “because [she] wants to do it.” (Frankfurt 2002: 27) If Beth was able to do otherwise in the circumstances than attempt to kill George only if she was able to show mercy, then, because her new system of values left no room for mercy, she was not able to do otherwise than attempt to kill George. When Beth falls asleep at the end of her bad day, the manipulators undo everything they had done to her. When she awakes the next day, she is just as sweet as ever (Mele 2006: 171–172).<sup>4</sup>

Mele labels manipulation cases like Beth’s “radical reversals,” and he contrasts them with the type of manipulation case that Watson apparently had in mind, in which a powerful being is an “original designer” of a world, or at least a deterministic sequence that includes an agent’s life from beginning to the act in question.<sup>5</sup>

Radical reversals are important in the present context because they pose a formidable challenge to internalist accounts of moral responsibility like Frankfurt’s. On the one hand, Beth seems clearly *not* morally responsible for killing George. On the other hand, Frankfurt’s account implies that she is.<sup>6</sup> And the same goes for other

---

<sup>4</sup> As an anonymous reviewer points out, Mele’s description of the case suggests that Beth lacks the ability to do otherwise than kill George, and while this would not matter to some compatibilists (namely semicompatibilists), it may make a difference to others, namely those who think that an agent is (directly) morally responsible for something only if the agent could have done otherwise. For compatibilists of the latter stripe, we may modify the case, stipulating that Beth is causally determined to kill George (because of her manipulation) but nevertheless retained that ability to do otherwise in any compatibilist-friendly sense one prefers.

<sup>5</sup> For more on “original-design” cases, see Mele’s discussion of his “zygote argument” in Mele (2006: 184–195; 2016: 71–72; and 2019: 83–84).

<sup>6</sup> Moreover, on his account, she is just as morally responsible for killing George as Chuck is when he commits the same crime. Here I disagree with Frankfurt, for reasons that will become clear in the following sections of the paper.

internalist accounts (like Watson's). As I mentioned above, it is in response to worries about manipulation that Frankfurt and Watson make the claim that compatibilism requires internalism.

Mele interprets Frankfurt's and Watson's claims from the previous section as making certain no-difference claims (Mele 2016: 81).<sup>7</sup> In particular, Mele sees Frankfurt and Watson as claiming that manipulation cases (including radical reversals) are not relevantly different from ordinary deterministic causation. Hence Frankfurt's claim that "[w]e are inevitably fashioned and sustained, after all, by circumstances over which we have no control" (Frankfurt 2002: 28). What matters for moral responsibility, then, are only the time-slice properties possessed by the agent at the time of the action (her psychological structure at that time), which is to say that internalism is true.

Mele criticizes the approaches of Frankfurt and Watson by attempting to show that there are relevant differences between radical reversals and ordinary deterministic determination. In particular, he argues that there are pairs of cases (which we can assume are deterministic) that generate asymmetric responses, where only one of those cases features radical-reversal-type manipulation. For example, return to the case of Chuck and Beth, and suppose that both kill someone named George. It would seem that Chuck is morally responsible for killing George, but that Beth isn't, despite the only difference between them being a difference in history (they have the same time-slice properties). Crucially, whereas Beth was "fashioned and sustained, after all, by circumstances over which [she had] no control," Chuck fashioned himself in a way that Beth did not, and this difference appears to explain the asymmetry in our moral responsibility judgments, which counts against internalism.

In the remainder of this paper, I will argue that there is a sense in which Mele is right and a sense in which he is wrong. My argument depends on an important distinction to which I devote the next section.

#### 4 The *Fact of Responsibility* and the *Degree of Responsibility*

Up to this point, I and the authors I have cited have been talking about *whether* agents are morally responsible in various conditions. But we can distinguish this question—whether an agent is morally responsible—from a distinct one about the

<sup>7</sup> This is one of two of what Mele (2019: 95) calls "weak branches" of the line of thought expressed by Frankfurt, Watson, and also Double (1991). The other branch is what Double seems to have in mind when he says that "the internalistic view is implicit in compatibilism" and that "compatibilism has not a chance of plausibility without [internalism], since otherwise the incompatibilist abhorrence of determinism will destroy it" (Double 1991: 56–57), which Mele interprets as "the idea that if manipulation of the sort involved in my radical reversal stories were to get an agent off the hook, it would do so only if it includes deterministic causation of crucial psychological events or states, in which case determinism would be the real culprit" (Mele 2016: 81). But, as Mele persuasively argues, there can be parallel cases of *indeterministic* manipulation that produce the same result, so this branch of the line of thought that compatibilism requires internalism is unsuccessful. For Mele's argument, see Mele (2016: 75–76) and the works cited there.

how morally responsible an agent is.<sup>8</sup> This is the distinction between the *fact* of an agent's moral responsibility and the *degree* of the agent's moral responsibility. Take, for example, an agent, S, who is blameworthy for X. Even given the fact that S is blameworthy for X, we might ask *how blameworthy* S is for X. In other words, there is a threshold for moral responsibility, but even among agents who meet the threshold conditions for moral responsibility we can ask about gradations of moral responsibility. If two agents perform the same type of action and are both morally responsible for doing so, it could be that one is *more morally responsible* than the other.<sup>9</sup>

To see this distinction in action, consider the first action for which you were morally responsible. If you're like me, you probably do not remember the first action for which you were morally responsible. Even so, it probably occurred sometime during your childhood, perhaps when you were four or five years old.<sup>10</sup> By stipulation, you met the threshold conditions of moral responsibility for this action, as it was the first action for which you were morally responsible. But presumably you were not *as* morally responsible for that action as you are for some action that you are morally responsible for performing today. What has changed? Many things have changed, to be sure, but here are a couple of differences that I take to be especially salient: first, you have a better *understanding* of morality, of the consequences of your actions, etc.; second, you have more *control* over your conduct, and over the character from which it stems. In other words, there are *epistemic* differences and *metaphysical* (or *control*) differences. Going back to Aristotle, many philosophers have taken there to be some epistemic condition and some control condition on moral responsibility. But how much we know and how much control we have can come in degrees, and arguably two agents may both meet the threshold conditions for moral responsibility and yet differ in degree of moral responsibility because of differences in the degree to which they understand or control their conduct.

One factor that affects an agent's degree of control over her actions is the degree to which it is a matter of luck that she performs the action. As Christopher Evan Franklin puts the point, "there seems to be an inverse relation between luck and control: the more an action is subject to luck, the less it is under our control, and the more an action is under our control, the less it is subject to luck" (Franklin 2011: 200). Additionally, the degree to which an agent's performing an action is a matter of luck depends on the degree to which the character from which her action stemmed is a matter of luck. Luck in having the character one has falls under the category of *constitutive luck*.<sup>11</sup> For example, return to the cases of Beth and Chuck. As I said

<sup>8</sup> Cf. Fischer (1985: 256), who distinguishes between the *content* of moral responsibility (what someone is morally responsible *for*) and the *extent* (or *degree*) of moral responsibility. Zimmerman (2002) uses this distinction in an attempt to solve certain problems of moral luck.

<sup>9</sup> I take moral responsibility itself to come in degrees, but some theorists, such as Fischer and Ravizza (1998), take moral responsibility to be an on/off and not a scalar concept. Even so, Fischer and Ravizza admit that praiseworthiness and blameworthiness may admit of degrees. My basic response to Mele will work equally well on this alternative conceptual framework, but I stick with my own preferred framework for simplicity's sake.

<sup>10</sup> For discussion of little agents, see Mele (2006: 129–133).

<sup>11</sup> This term was introduced by Nagel (1979: 28).

near the end of the previous section, Chuck fashioned himself in a way that Beth did not (Beth was “fashioned and sustained, after all, by circumstances over which [she had] no control”). Given Chuck’s history, it is not as much a matter of luck that he has the character that leads to his killing George as it is a matter of luck that Beth has the character that leads to her killing George—in fact, Beth had no control over her having that character and so is entirely constitutively (un)lucky. And, as I said before, this difference appears to explain the asymmetry in our moral responsibility judgments about the two agents.<sup>12</sup>

Given the distinction introduced in this section, however, we must ask whether Beth and Chuck differ with respect to the *fact* of moral responsibility, the *degree* of moral responsibility, or both. Since they differ in their degree of control, I believe that they differ in degree of moral responsibility. But this is consistent with internalism, according to which the *fact* of moral responsibility is a time-slice (or structural) matter. To count as an externalist position, one must go beyond this and claim in addition that Beth and Chuck differ with respect to the *fact* of moral responsibility.

It is worth pausing at this point to consider an objection. On my view, Beth is less morally responsible than is Chuck, and this is because Beth has less control than does Chuck. But, one might object, “less control” is ambiguous between “lower degree of control,” on the one hand, and “control over fewer things,” on the other, and it may be that Beth is less morally responsible than Chuck because she has control over fewer things, not because she has a lower degree of control.<sup>13</sup> This is a very important objection, and I address it in more detail elsewhere.<sup>14</sup> I agree that Chuck had control over more things than did Beth, but to say that the difference in their degree of moral responsibility reduces to the difference in the scope of their control would be, I think, to misunderstand the way in which constitutive luck mitigates moral responsibility. Insofar as Beth did not have control over the events that led to her having a bad character, her having the bad character that leads to killing George is entirely a matter of (constitutive) luck. By contrast, Chuck’s having the bad character that leads to his killing George is not entirely a matter of (constitutive) luck. And as long as we take luck and control to be inversely related, as I suggested above, attending to Beth’s constitutive luck should lead us to take her control over killing George (and thus her moral responsibility for doing so) to be lesser in degree compared to Chuck’s.

<sup>12</sup> As an anonymous reviewer has encouraged me to highlight, I am offering a new response to cases of radical reversal. Whereas Frankfurt would say that Beth is just as morally responsible as Chuck, and whereas Mele would say that Beth is not morally responsible at all, my view is that Beth is a little bit morally responsible but not nearly as morally responsible as Chuck.

<sup>13</sup> Thanks to Neal Tognazzini for raising this objection. For more on the distinction between scope/content of moral responsibility (and control), on the one hand, and degree of moral responsibility (and control), on the other, see Fischer (1985) and Zimmerman (2002). And for an attempt to use this distinction in defense of internalist (structuralist) compatibilist views, see McKenna (2012).

<sup>14</sup> See Cyr (Forthcoming).



## 5 An Argument for Internalism (about the *Fact* of Responsibility)

I will now argue that, so long as one is a compatibilist, one should reject the view that Beth and Chuck differ with respect to the *fact* of moral responsibility. And since there is nothing special about Beth and Chuck, the argument may be extended to any pair of manipulated and non-manipulated agents who satisfy the same internalist conditions at the time of action and so constitutes an argument for the claim that compatibilism requires internalism about the fact of moral responsibility, even if this may be combined with an *externalist* position about the degree of moral responsibility (which, as it happens, I endorse). Here is the argument:

1. Agents who are entirely constitutively lucky can be morally responsible for what they do.
2. There is no relevant difference between agents like Beth and agents who are morally responsible for actions that stem from characters with respect to which they are entirely constitutively lucky.
3. Agents like Beth can be morally responsible for what they do, despite satisfying only internalist conditions on moral responsibility (i.e., failing to satisfy additional, externalist conditions), which is to say that externalism is false.<sup>15</sup>

It was this line of reasoning that convinced me—previously an externalist about moral responsibility—to accept internalism instead. I will now explain why compatibilists are committed to the truth of the first premise and why, I think, they should accept the truth of the second as well and thus endorse internalism.<sup>16</sup>

Compatibilists believe that agents like us—agents who began to exist—can be morally responsible despite being causally determined. But if an agent who began to exist can be morally responsible for her actions, then there must be some *first* action for which she is morally responsible (hereafter “first morally responsible action”). But if an action is an agent’s *first* morally responsible action, then she is not morally responsible for the character from which her action stems, and so she is entirely constitutively lucky at the time of her first morally responsible action. There can be cases, of course, in which an agent’s constitutive luck is mitigated over time, as seems to be the case for Chuck, who actively contributes to the formation of his character. But this is not the case for agents’ *first* morally responsible actions, so compatibilists must accept that we can be morally responsible while entirely constitutively lucky, which is to say that they are committed to the truth of the first premise.

<sup>15</sup> This argument is based on the main argument of Cyr (Forthcoming).

<sup>16</sup> While I do say that compatibilists *should* accept the second premise (along with the first) and thus endorse internalism, I do not mean to suggest that my thesis is really only that compatibilists should be internalists. What I did mean to suggest is that, whereas the first premise follows straightforwardly from compatibilism, there is an independent argument for the second premise, which I provide below and believe is sound. Unless it is unsound, compatibilists must be internalists. Thanks to an anonymous reviewer for encouraging me to clarify.

The second premise says that there is no relevant difference between agents like Beth and agents who are morally responsible for actions that stem from characters with respect to which they are entirely constitutively lucky. To see why this is true, consider first a crucial similarity between Beth and an agent who performs her first morally responsible action: both act from a character and set of values over which they exercised no control. While Beth may not be *entirely* constitutively lucky, since the manipulators need not have completely overhauled her character, when it comes to the action of killing George, the relevant aspects of Beth's constitution (her new values and character which make it psychologically possible for Beth to kill George) were not under her control. Moreover, whatever can be said in favor of the moral responsibility of the agent performing her first morally responsible action can be said about Beth as well. Both are mentally healthy, rationally competent, aware of the moral significance of the action, etc., which is to say that both can satisfy your favorite internalist criteria for morally responsible action. Finally, any consideration that might exonerate Beth for killing George would seem also to exonerate agents at the time of their first morally responsible actions. Neither Beth nor these other agents is (at least yet) morally responsible for endorsing their character. And, going back to a point that Frankfurt highlights, both have been "given" characters over which they had no say. Beth's was intentionally produced, whereas the others' are the result of "natural forces," but both "are inevitably fashioned and sustained, after all, by circumstances over which [they] have no control" (Frankfurt 2002: 28).

I have argued that compatibilists must accept that agents who are entirely constitutively lucky can be morally responsible for what they do and that Beth is not relevantly different from such agents and so can be morally responsible as well. If this argument is sound, then internalism is true, since the fact of whether or not an agent is morally responsible must not depend on an agent's history (or else Beth would be off the hook) but rather must depend only on time-slice features of the agent. In spelling out and defending this argument, I have tried to supply an argument that would make sense of some of Frankfurt's counterintuitive claims.<sup>17</sup> But whereas Frankfurt would apparently see *no* difference between manipulated agents (like Beth) and non-manipulated agents (like Chuck) who satisfy his internalist criteria for moral responsibility, I argued in the previous section that manipulated agents have less control over what they do than relevantly similar non-manipulated agents do, and arguably this difference in control grounds a difference in degree of moral responsibility. So while Frankfurt may be happy to bite the bullet and accept that there is no difference in moral responsibility between Beth and Chuck, it is open to internalists to soften the blow, admitting Beth's moral responsibility but only to a very limited degree.<sup>18</sup>

<sup>17</sup> The argument is in the same spirit as Watson's argument too, and I agree with Watson's claim that there is no relevant difference between ordinary determined agents and agents who are the product of super-powerful designers.

<sup>18</sup> As Neal Tognazzini points out, the implication that Beth's degree of moral responsibility is similar to that of a young child's makes my view sound much more like an externalist view than an internalist one. My view certainly differs from typical internalist views (like Frankfurt's) in that I take Beth's moral responsibility to be significantly mitigated by her manipulation, but the view is nevertheless an internalist one, given its commitment to the non-historicity of the fact of moral responsibility. Moreover, one advan-

Before concluding, it is worth noting that it follows, on my view, that there is an importance difference between agents who undergo “radical reversals,” on the one hand, and agents who are “originally designed,” on the other.<sup>19</sup> Suppose that we add to the case of Chuck that, years before undertaking his heart-hardening mission, Chuck was created by a goddess (in a deterministic world) who intended for him to develop into an agent who would undertake this mission and eventually kill George.<sup>20</sup> And recall that Beth does not contribute to the formation of the character from which she acts when she kills (another guy named) George. Whereas Mele’s view implies that Beth is not morally responsible for killing George but that, if compatibilism is true, Chuck *is* morally responsible for killing George, my view is that *both* are morally responsible for killing George, but Chuck is morally responsible to a much greater degree than is Beth. And the reason that Chuck is more morally responsible than Beth is because he had more control than she did (since he was less constitutively lucky than her).<sup>21</sup> Thus, I agree with Mele that there is an important difference between radical reversals and original designs, but we disagree about what exactly the difference is: for Mele, is a difference in the fact of moral responsibility; for me, it is a difference in degree of moral responsibility.<sup>22</sup>

## 6 Conclusion

In this paper, I have supplied a new argument for the claim that compatibilism requires internalism. The argument stems from the ideas expressed in some oft-quoted passages from Frankfurt and Watson, but given the distinction between *whether* an agent is morally responsibly, on the one hand, and *how* morally responsible the agent is, on the other, I have shown that internalists need not countenance

---

Footnote 18 (continued)

tage of my view over views like Frankfurt’s is that it is consistent with the soundness of the argument of this paper but also accommodates some of what is attractive about externalism (especially that Beth’s moral responsibility seems importantly different from Chuck’s).

<sup>19</sup> Note that Watson, in the quotation considered above, only discussed “designed” agents and not those who undergo mid-life reversals. Without more information, I’m not sure whether he’d side with Frankfurt or accept my asymmetric treatment of the two types of cases.

<sup>20</sup> Again, see Mele’s discussion of his “zygote argument” in Mele (2006: 184–195; 2016: 71–72; 2019: 83–84).

<sup>21</sup> On my view, there is no relevant difference between an original design scenario and an ordinary causally deterministic scenario. For an alternative view, according to which the effective intentions of the designer make the original design scenario relevantly different from ordinary determinism, see Waller (2014). It is worth noting that many compatibilists, including McKenna (2008), Fischer (2011), and Sartorio (2016) do not think the intentions of another agent makes a relevant difference.

<sup>22</sup> My view also differs from that of Barnes (2016), and one way to see the difference is by considering his case of Patty. Patty is similar to Beth in several key respects except that, instead of her change in character being due to manipulation, it is due to a “spontaneous neural evolution that is explainable in entirely naturalistic terms” (Barnes 2016: 2320). On Barnes’s view, Patty is clearly morally responsible, and presumably just as morally responsible as is Chuck (though this is not made explicit, and perhaps Barnes does not accept it). While my view would agree about Patty’s being morally responsible, it also implies that she is only morally responsible to a slight degree.

the counterintuitive claim that there is no difference in moral responsibility between non-manipulated agents and various manipulated agents. While compatibilists must be internalists about the *fact* of moral responsibility, it is open to compatibilists to be externalists about the *degree* of moral responsibility. Indeed, taking constitutive luck seriously should lead compatibilists in that direction, as the degree to which an agent is constitutively lucky is at least partly a historical matter.

**Acknowledgements** Thanks to Al Mele for discussing the main argument of this paper, and thanks to Gabriel De Marco, Neal Tognazzini, and two anonymous reviewers for comments on the paper.

## References

- Barnes, E.C. 2016. Character control and historical moral responsibility. *Philosophical Studies* 173: 2311–2331.
- Cyr, T. Forthcoming. Manipulation and constitutive luck. *Philosophical Studies*.
- Double, R. 1991. *The non-reality of free will*. Oxford: Oxford University Press.
- Fischer, J. M. 1985-6. Responsibility and failure. *Proceedings of the Aristotelian Society* 86: 251-270.
- Fischer, J.M. 2011. The Zygote Argument remixed. *Analysis* 71: 267–272.
- Fischer, J.M., and M. Ravizza. 1998. *Responsibility and control: A theory of moral responsibility*. Cambridge: Cambridge University Press.
- Frankfurt, H. 1988. *The importance of what we care about*. Cambridge: Cambridge University Press.
- Frankfurt, H. 2002. Reply to John Martin Fischer. In *Contours of Agency*, ed. S. Buss and L. Overton. Cambridge: The MIT Press.
- Franklin, C.E. 2011. Farewell to the luck (and *Mind*) argument. *Philosophical Studies* 156: 199–230.
- McKenna, M. 2008. A hard-line reply to Pereboom's four-case manipulation argument. *Philosophy and Phenomenological Research* 77: 142–159.
- McKenna, M. 2012. Moral responsibility, manipulation arguments, and history: Assessing the resilience of nonhistorical compatibilism. *Journal of Ethics* 16: 145–174.
- Mele, A. 1995. *Autonomous Agents*. Oxford: Oxford University Press.
- Mele, A. 2006. *Free Will and Luck*. Oxford: Oxford University Press.
- Mele, A. 2016. Moral responsibility: Radical reversals and original designs. *Journal of Ethics* 20: 69–82.
- Mele, A. 2019. *Manipulated Agents: A Window to Moral Responsibility*. New York: Oxford University Press.
- Nagel, T. 1979. *Moral luck*. In *Mortal Questions*. Cambridge: Cambridge University Press.
- Sartorio, C. 2016. *Causation and Free Will*. New York: Oxford University Press.
- Waller, R. 2014. The threat of effective intentions to moral responsibility in the zygote argument. *Philosophia* 42: 209–222.
- Watson, G. 1999. Soft libertarianism and hard incompatibilism. *Journal of Ethics* 3: 353–368.
- Zimmerman, M. 2002. Taking luck seriously. *Journal of Philosophy* 99: 553–576.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.