

Alfred R. Mele
Florida State University
amele@fsu.edu

Original Scientific Paper
UDC 123.1
17.021.1

Received: August 14, 2024

Revised: September 19, 2024

Accepted: September 23, 2024



MORAL RESPONSIBILITY: AGENTS' HISTORIES AND LITTLE AGENTS

Abstract

This paper addresses a pair of issues about moral responsibility. One is whether “externalists” should claim that agents must have agential histories of a certain kind in order to be morally responsible for actions they perform or should instead, in light of certain stories about imaginary full-blown agents who have no agential history – that is, have never performed an action (because they just now came to exist) – stop short of that and contend that agents must lack an agential history of a certain kind in order to be morally responsible for something they do. (Such imaginary beings have come to be called “instant agents”.) The other issue concerns the earliest actions for which ordinary human agents are morally responsible. How does an ordinary child’s moral responsibility get off the ground, as it were? Connections between these two issues are explored, as is the bearing of some kinds of manipulation on moral responsibility. It is assumed that only beings who are capable of acting freely are capable of being morally responsible for actions they perform.

Keywords: agential histories · children · free will · instant agents · manipulation · moral responsibility

In previous work (including Mele 1995, 2006, 2019), I have argued for what has been called *historicism* or *externalism* about moral responsibility, a view according to which agents’ histories have an important bearing of a kind to be described shortly on whether or not they are morally responsible for an action of theirs. My strategy included (among other things) presenting counterexamples to opposing internalist views and, based partly on those counterexamples, arguing for a necessary condition for moral responsibility for actions – specifically, *direct* moral responsibility – that entails externalism.

In this paper, I take up a pair of issues that continue to be topics of lively discussion. One is whether externalists should claim that agents must *have* agential histories of a certain kind in order to be morally responsible

for actions they perform or should instead, in light of certain stories about imaginary full-blown agents who have no agential history – that is, have never performed an action (because they just now came to exist) – stop short of that and contend that agents must *lack* an agential history of a certain kind in order to be morally responsible for something they do.¹ The issue here is about the relative merits of *positive* and *negative* proposed historical requirements for moral responsibility for actions. All positive historical requirements (as that expression is understood here) entail, to put it simply, that no possible agent can be morally responsible for its first action, and negative historical requirements do not entail this. The other issue to be explored concerns the earliest actions for which ordinary human agents are morally responsible.² Connections between these two issues will emerge in what follows.

1. Background

The following is a relatively concise statement of externalism about direct moral responsibility:

agents sometimes are [directly] morally responsible for *A* partly because of how they came to be in the internal condition that issues in their *A*-ing; and, more specifically, in these cases, there is another possible way of having come to be in that internal condition such that if they had come to be in that condition in that way, then, holding everything else fixed (except what is entailed by the difference, of course), including the fact that they *A*-ed, they would not have been morally responsible for *A*. (Mele 2019, pp. 13–14)

Brief comments on a pair of technical terms in this passage are in order. First, as I understand *direct* moral responsibility, “an agent is *directly* morally responsible for *A*-ing when and only when he is morally responsible for *A*-ing and that moral responsibility is not wholly inherited from his moral responsibility for other things” (Mele 2019, p. 11). Second, an agent’s *internal condition* at a time, as I understand it, is “something specified by the collection of all psychological truths about the agent at the time that are silent on how he came to be as he is at that time” (Mele 2019, p. 6).

To give readers a feel for claims of the kind externalists oppose, I quote the following from a pair of papers by Harry Frankfurt:

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

1 For two recent discussions, see Fischer 2021 and McKenna 2021.

2 For two recent discussions, see Cyr 2020 and McKenna 2021.

the questions of whether he performs the actions freely or is morally responsible for performing them. (Frankfurt 1988, p. 54)

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, p. 27)

Frankfurt is claiming here that as long as an agent is in a certain internal condition when he *A*-s, he is morally responsible for *A*-ing, no matter how he came to be in that condition. Gary Watson is another frequently cited internalist about direct moral responsibility for actions (see Watson 1999).

One more bit of background will be useful before I forge ahead. In various thought experiments, including some discussed below, I use the term “values” as a kind of shorthand. As I understand the term “values” in Mele 1995 (and subsequently), “*S* at least *thinly values X* at a time if and only if at that time *S* both has a positive motivational attitude toward *X* and believes *X* to be good” (p. 116). When values are understood as psychological states, I take them to have both of these dimensions by definition. As I observe in Mele 1995 and elsewhere, this account of thinly valuing and the corresponding thin account of values are not meant to be contributions to the theories of valuing and values; their purpose is simply to make my meaning clear.

2. Historical Requirements for Moral Responsibility

I turn now to the relative merits of positive and negative proposed historical requirements for moral responsibility for actions. Some far-fetched scenarios have motivated me to be cautious when formulating an externalist necessary condition for morally responsible action. One such scenario, from Mele 1995, features an agent, Athena, who “magically comes into existence with a wealth of beliefs, desires, and values in place” (p. 172). Athena “has a robust ability to shed each of her pro-attitudes – an analogue of the ability that we have to refrain from acting on resistible desires” (p. 173). “Other things being equal,” I argued, “there is no bar to Athena’s autonomously possessing her pro-attitudes” (p. 173). If Athena is a possible being and if her value-shedding ability is such that she can shed values instantaneously, she may autonomously possess values throughout a stretch of time that begins when she begins to exist. And if she autonomously possesses the values she has when she comes into existence, she would seem to be capable of being morally responsible for

actions performed on the basis of those values at that time. Obviously, when she comes into existence, Athena has no history as an agent – that is, she has not yet performed any actions.

I myself have never been prepared to claim that Athena – a being with a sophisticated psychological profile who can act in the first moment of its existence – is indeed a possible being. And, as I have pointed out, if externalists about mental content like Donald Davidson are correct, Athena is not a possible being (Mele 1995, p. 172). On a view like Davidson's, because of the way Athena comes to exist, she has no thoughts at all at this time and therefore is not yet in a position to act (see Davidson on Swampman, 1987, p. 444). But Davidson's view is controversial; and for all I know, Athena is a possible being. My claim is that if Athena is a possible being, she falsifies the assertion that any possible agent must have an agential history in order to be morally responsible for an action it performs.

Beings of another kind are relevant here. I call them “minutelings” (Mele 2019, pp. 55–61). Here is the basic idea:

Cloned human bodies are kept unconscious for years. Their muscles get a lot of daily exercise as the bodies grow and mature. For each body, a complete psychological profile of some actual human being or other is created, including a whole raft of pseudo-memories that match the human being's actual memories. At a certain point, the body and psychological profile are merged. And now, for the first time, a conscious being is located where the body is. Given the skills and technology of the people in charge, the being can live no more than a minute. (pp. 55–56)

I give a copy of the psychological profile of a normal, reasons responsive, morally responsible agent to a minuteling and place both the minuteling (at the moment at which the merger occurs) and the normal agent in exactly similar situations calling for a decision to be made right then about a morally significant matter (pp. 56–57).

I do not take a stand on whether the minuteling is or is not morally responsible for the decision he makes. I say that I am agnostic about that (Mele 2019, p. 60). Now, even if Davidson is right about a being like Athena, the way in which a minuteling gets his psychological profile – that is, the causal route – allows him to have beliefs, desires, values, and the like. If a particular minuteling is morally responsible for the first action he performs after the merger, does that falsify the claim that any possible agent must have an agential history in order to be morally responsible for an action it performs? (The first action may be a decision. Like many philosophers, I count deciding to do something as a mental action.)

Matters are not clear cut. Are the minuteling and the person whose psychological profile he was given two different persons (as I believe)? Or are

they, at the moment of the merger, the same person in two different places?³ If the latter, then the minuteling's first action actually was performed long ago (that is, long before his minuteling "stage"). If the former, a minuteling might undermine the claim that any possible agent must have a history as an agent in order to be morally responsible for an action it performs.

"Instant agents" is David Zimmerman's term for agents "who spring full-blown into existence. . . . Mele's 'Athena' and Davidson's 'swampman' are vivid examples" (1999, p. 252). If, in the scenarios at issue, Athena fails to be an instant agent because she fails to be an agent at the moment she comes into existence and minutelings fail to be instant agents because they do not spring full-blown into existence, then perhaps externalists should opt for a positive historical requirement for being morally responsible for an action. But would we learn much about moral responsibility from these alleged failures? Or would we instead learn something about what it takes to be an agent and something about personal identity?

These questions merit some attention. One path into this topic goes through what I have argued to be a necessary condition for being morally responsible for an action. I have argued for this largely on the basis of various alleged counterexamples to internalist views of moral responsibility. Consider the following:

NFMR. An agent does not freely *A* and is not morally responsible for *A*-ing if the following is true: (1) for years and until manipulators got their hands on him, his system of values was such as to preclude his acquiring even a desire to perform an action of type *A*, much less an intention to perform an action of that type; (2) he was morally responsible for having a long-standing system of values with that property; (3) by means of very recent manipulation to which he did not consent and for which he is not morally responsible, his system of values was suddenly and radically transformed in such a way as to render *A*-ing attractive to him during *t*; and (4) the transformation ensures either (*a*) that although he is able during *t* intentionally to do otherwise than *A* during *t*, the only values that contribute to that ability are products of the very recent manipulation and are radically unlike any of his erased values (in content or in strength) or (*b*) that, owing to his new values, he has at least a Luther-style inability during *t* intentionally to do otherwise than *A* during *t*. (Mele 2019, pp. 66–67)

The expression "Luther-style 'inability'" refers to the sort of (alleged) inability Daniel Dennett seems to have in mind in his influential discussion of Martin Luther's remark that he could not do otherwise than register his protest (Dennett 1984, p. 183; for discussion, see Mele 2019, pp. 66–67).

3 Evan Jones raised this question in discussion.

If NFMR is true, then a *necessary* condition of an agent's being morally responsible for A-ing is that he not satisfy the conjunction of conditions 1 through 4. And if that is so, externalism is true; this proposed necessary condition has historical elements.⁴ The proposed necessary condition does not assert that an agent must *have* an agential history of some kind or other in order to be morally responsible for an action; it does not assert that the agent must have any history as an agent at all. It is compatible with the existence of an instant agent who is morally responsible for something he does in the first moment of his existence. The alleged necessary condition for moral responsibility is, then, a *negative* historical condition. What it requires is the *lack* of a history of a certain kind.

For someone so inclined, turning this negative condition into a positive one is easy. Consider the following: (POS) An agent is morally responsible for A-ing only if he has a history as an agent and does not satisfy the conjunction of conditions 1 through 4. Regarding the nature of moral responsibility in particular, this alleged positive historical requirement adds very little to the alleged negative one. What it adds is the claim that one cannot be morally responsible for an action one performs unless one has already acted at least once. And readers who believe that agents like Athena or instant-agent minutelings are possible beings who might be capable of being morally responsible for their first actions will not see themselves as losing anything important in rejecting this modest addition to the negative historical requirement.

Attention to a considerably more robust alleged positive historical requirement for moral responsibility than POS will prove instructive. John Fischer and Mark Ravizza contend that being (directly) morally responsible for an action one performs depends on the "the agent's 'ownership' of the mechanism" that issues in the action and that agents "*make* certain kinds of mechanisms *their own*" – that is gain *ownership* of those mechanisms – "by *taking responsibility* for them" (1998, p. 241). Taking responsibility, on their view, "is a matter of having certain (dispositional) beliefs about oneself (and having acquired those beliefs in appropriate ways)" (p. 243). Fischer and Ravizza describe the beliefs in question in terms of how the agent "views himself":

First, an agent must view himself – when acting from certain mechanisms – as an agent; he must see that certain upshots in the world are the results of his choices and actions. Second, an agent must view himself as an apt target for the reactive attitudes. . . . Finally, the cluster of beliefs specified by the first two conditions must be based, in an appropriate way, on the individual's evidence. (p. 238)⁵

4 I assume that the agent's moral responsibility for his system of values is at least primarily due to actions he performed.

5 For two versions of second condition, see Fischer and Ravizza 1998, p. 238.

Fischer and Ravizza write: "On our account, taking responsibility is a genuinely historical notion" (1998, p. 238). If an agent's having acquired the beliefs at issue "in appropriate ways" (p. 243) is meant to entail his having acquired those beliefs partly in response to consequences of past actions of his, then taking responsibility, as they conceive of it, has a historical dimension.

How are internalists about moral responsibility for actions likely to respond to the claim that an agent's being morally responsible for an action he performs depends on his "taking responsibility," in Fischer and Ravizza's sense of that expression, for the mechanism that issued in that action? Imagine, if you can, an instant agent (perhaps a minuteling) who comes into existence with a view of himself as someone whose "choices and actions" have "upshots in the world" and as "an apt target for the reactive attitudes" (Fischer and Ravizza 1998, p. 238). An internalist who takes such an agent to be possible and who agrees with Fischer and Ravizza that such a view of oneself is required for being morally responsible for what one does will contend that although that self-conception is required for moral responsibility, it does not matter how the self-conception came to be present. At this point, Fischer and Ravizza need either an argument that such instant agents are impossible or a plausible explanation of why, even if they are possible, the self-conception they are assumed to have is not equipped to do the work called for in the present connection unless it was acquired partly in response to consequences of past actions the agent performed.⁶

Fischer and Ravizza report their belief that "two sorts of cases" make it reasonable to contend that "moral responsibility is indeed a genuinely historical notion" (1998, p. 195). The first kind of case concerns indirect moral responsibility, as when a drunk driver's moral responsibility for accidentally killing a pedestrian traces back to an action or actions for which he was directly morally responsible. Since my concern here is direct moral responsibility, I set cases of indirect moral responsibility aside. The second kind of case features manipulation of various kinds (pp. 196–201): for example, hypnosis, brainwashing, or "direct stimulation of the brain" (p. 197). But although carefully crafted cases of this kind may directly support the judgment that agents must *lack* histories of certain kinds in order to be morally responsible for an action (that is, histories that include certain kinds of manipulation), more would need to be done to support the more demanding claim that *having* an agential history (of a certain kind) is required for moral responsibility for actions.

My preferred way of motivating externalism about direct moral responsibility for actions is by way of stories about heavy-duty manipulation in which an agent who worked hard to shape his character and succeeded at that is given a radically reversed system of values overnight (without his

6 Fischer reports that he is willing to concede that instant agents are possible for the sake of argument (2021, p. 260).

consent) and acts on the basis of his new values the next morning (Mele 2019). For example, an exceptionally nice person who worked hard to make herself that way is given the values of a ruthless killer. It is easy to feel the pull of the verdict that the agent is not morally responsible for the action(s) at issue (even if, in the end, one rejects the verdict). To be sure, such stories highlight the agent's past behavior in addition to heavy-duty manipulation, but that is entirely consistent with my stopping at a negative historical requirement for direct moral responsibility for actions. Once again, what prevents me from moving past that to a positive requirement in terms of agential histories is the thought that instant agents of certain kinds might be counterexamples to the thesis that no possible agent can be morally responsible for its first action. This thesis is entailed by any positive historical requirement of the sort at issue, as I have mentioned.

Compare a “radical reversal” story of the kind just mentioned with a story about an agent, Phil, who, for philosophical reasons, has for years rejected the proposition that “it is *prima facie* plausible that he is an appropriate target for the reactive attitudes” (Fischer and Ravizza 1998, p. 238).⁷ Phil believes, on philosophical grounds, that no one is an appropriate target for the reactive attitudes. Like friends of mine who share this view with him, Phil is a nice person who donates time and money to worthy causes, helps neighbors in need of assistance, and so on. Overnight, he is given a new view of himself (and others) – a view according to which he had been misled by faulty philosophical reasoning and is in fact an appropriate target for the reactive attitudes when acting from various mechanisms. The new view is implanted by means of new-wave brain surgery that does not in any way draw on Phil's past experiences as an agent. The next morning, a Saturday, Phil spends several hours working on a Habitat for Humanity house, as he often does on weekends. And this time he takes himself to deserve some positive moral credit for his good deed. Getting a majority of readers to feel the pull of the verdict that Phil is not morally responsible for his good deed would not be an easy task. Even readers who agree that Phil must believe that “it is *prima facie* plausible that he is an appropriate target for the reactive attitudes” in order to be morally responsible for his good deed may balk at the further claim that his being morally responsible for that deed also depends on his having acquired that belief at least partly in response to past actions of his. (Recall that the historical dimension of Fischer and Ravizza's view is found in the third condition of their account of taking responsibility, a condition about the basis of “the cluster of beliefs specified by the first two conditions” [p. 238].)

7 At this time in his life, because he rejects the proposition about *prima facie* plausibility, Phil fails to satisfy an alleged requirement for taking responsibility that Fischer and Ravizza propose for agents who have engaged in “philosophical reflection on the relationship between causal determinism and the appropriateness of the reactive attitudes” (p. 238). For reservations about the alleged requirement, see Mele 2006, pp. 146–55.

I certainly do not mean to suggest that Fischer and Ravizza's route to an alleged positive historical requirement for moral responsibility for actions is the only candidate for a route to such a requirement. Elsewhere, I have argued against an alternative route suggested in McKenna 2012 (see Mele 2019, pp. 49–61). My aim in the preceding discussion of taking responsibility has been to identify challenges faced by a prominent positive externalist view of what is required for direct moral responsibility for actions. As things stand now, I believe that the safest course for opponents of externalism is to stop at a very plausible alleged negative historical requirement. If such an alleged requirement is, in fact, a requirement, internalism is false – in which case, externalism is true.

3. Little Agents

I turn now to the second issue I said I would discuss, the earliest actions for which ordinary human agents are morally responsible. Assume for the sake of argument that moral responsibility for actions is widespread in the actual world. In our world, children act – and act intentionally – before they are in any position to be morally responsible for actions of theirs. Our children have histories as agents before they become morally responsible agents. Consider a normal child, Mickey, a very nice little boy who arguably is performing the first action for which he is morally responsible, a kind action toward a playmate.⁸ Suppose that although, for quite a while, Mickey has viewed himself as someone whose “choices and actions” have “upshots in the world” (Fischer and Ravizza 1998, p. 238), he very recently crossed the threshold for viewing himself as an apt target for the reactive attitudes. Suppose also that he acquired the beliefs at issue in what Fischer and Ravizza would deem “appropriate ways” (p. 243) and that he satisfies their well-known conditions for moderate reasons responsiveness when he performs his kind action (pp. 89–90).

With these suppositions in place, along with the assumption that moral responsibility for actions is widespread in our world, many readers may see Mickey as having some moral responsibility for his kind action. Now compare Mickey at the time of action with two other agents who behave as Mickey does at that time. One is an instant agent (perhaps a minuteling) with Mickey's psychological profile, including his conception of himself as an apt target for the reactive attitudes and as a person whose actions have consequences. And the other, Mike, is a little rascal who would definitely have behaved unkindly toward his playmate at the time if skilled manipulators had not, while he slept last night, replaced his rascally collection of values with values that match

8 I am not assuming the absence of a gray area between actions for which a little agent has no moral responsibility and actions for which he has some. What is (as I put it) arguably the first action for which a little agent has some moral responsibility may also be argued to be in the gray area instead. To simplify discussion, I forgo discussing vagueness.

Mickey's. (Mike also was given in the process a view of himself as an apt target for the reactive attitudes. He already viewed himself as someone whose actions have upshots in the world.)

As in the case of "adult" instant agents with values that match those of ordinary morally responsible adult agents, I find that I do not lean one way or the other regarding the question of whether or not Mickey's instant-agent counterpart is morally responsible for the action he performs. I am agnostic about that as I am about the adult instant agents (Mele 2019, p. 60). I should also point out that even if this counterpart is morally responsible for the action at issue, that does not falsify my externalist claim that a necessary condition of an agent's being morally responsible for *A*-ing is that he not satisfy the conjunction of conditions 1 through 4 of NFMR. After all, the counterpart does not satisfy this collection of conditions; indeed, he does not satisfy any of these four conditions.

What about Mike, the manipulated child? I am strongly disinclined to count him as morally responsible for his good behavior at the time at issue. Obviously, I cannot appeal to NFMR in an attempt to account for Mike's lack of moral responsibility for his first post-manipulation action; like Mickey's instant-agent counterpart, Mike does not satisfy NFMR's conditions. So why do I view him as not being morally responsible for that action? Because the values from which that action flowed – values with a very different location on the nice-versus-nasty spectrum from the rascally values that were erased – were implanted by manipulators (without Mike's consent) and because there would have been no chance of his performing the good deed were it not for the manipulation. (This is not to say that agents are never morally responsible for actions that flow from implanted values; see Mele 2019, pp. 36–38.)

Some readers will object that, given that we are considering what are supposed to be the *first* actions for which Mickey and Mike might be morally responsible, it cannot matter how these two little agents came to be in the internal condition they were in at the time of action. After all, by hypothesis, they were not morally responsible for doing anything that contributed to their being in that internal condition. In this vein, regarding a similar pair of little agents, Michael McKenna writes: "It seems to me that for both Tony and Tara, it is just dumb luck what values they start with, because either way it is beyond their control. How they acquired them has no bearing on whether when they act on them they are a teeny bit free and responsible as their careers as moral agents initially get up and running" (2021, p. 287; also see Cyr 2020).

That this line of thought may be too quick is easier to see from another angle. Imagine a scenario in which the manipulators give Mickey the little rascal's collection of values overnight, and, as a consequence, Mickey gleefully performs some unkind action the next morning.⁹ A predictable reaction is

9 For a scenario of this kind, see McKenna 2021, pp. 287–88.

that it seems unfair to deem Mickey blameworthy for this action, despite the fact that his pre-manipulation collection of values was not at all shaped by any behavior of his for which he was morally responsible. Why is that?

Even though, by hypothesis, Mickey had not reached the threshold for moral responsibility prior to the day at issue, he presumably had learned from some of the consequences of his behavior and had modified some of his behavior accordingly. This would amount to his having had some control over his collection of values (in my thin sense of "values," of course). The control he had at these times was not, by hypothesis, moral-responsibility-level control. Even so, a collection of considerations speaks in favor of the claim that deeming Mickey blameworthy for his bad behavior at the time would be unfair. In addition to his modest contributions to the constitution of his pre-manipulation values, there is the fact that those values had a very different location on the nice-versus-nasty spectrum from the rascally values implanted by the manipulators and the fact that, were it not for the manipulators, there would have been little chance of Mickey's performing a nasty action at the time.

Is someone who appeals in this way to Mickey's pre-manipulation behavior (together with the two facts about manipulation just mentioned) thereby committed to the thesis that there is a positive historical requirement for direct moral responsibility for actions? No. One can make this appeal while leaving it open that some instant agents may be morally responsible for their earliest actions.

McKenna floats an idea that he dubs "Little Agent Internalism":

Unlike instant agents who come with fully formed moral personalities, for agents of the sort we human beings are, who emerge over time and so have a history of gradually acquiring the agential abilities and values on the basis of which they might engage in exercises of moral agency, a version of internalism suffices – *but just for the little agents*. It does not matter how these agents acquire their initial values. Rather, what matters is their content. Some contents are so objectionable that they should be regarded as impairing rather than enabling moral agency. (2021, p. 288)

I agree that some collections of values might impair moral agency. They might, for example, be incompatible with an agent's being able to apprehend moral reasons. But, as I just reported, I am inclined to believe that the way in which Mickey acquired his rascally values does "matter"; it seems to me to get him off the hook for the post-manipulation unkind action he performed. (Readers are encouraged to make their own judgments about the scenario at issue.)

McKenna's Little Agent Internalism is an instance of what is sometimes referred to as a *hybrid* view about direct moral responsibility for actions, insofar as it apparently is internalist about some scenarios and externalist about

others. But it should be observed that the view will be counted by proponents of NFMR as a species of externalism. Ordinary little agents performing the first actions for which they are morally responsible certainly are not exceptions to the *externalist* claim that a necessary condition of an agent's being morally responsible for A-ing is that he not satisfy the conjunction of conditions 1 through 4 of NFMR; they do not falsify this perfectly general externalist claim. The same goes for Ishtiyaque Haji's "hybrid view," according to which direct moral responsibility "does not require that one have a past, but should one have a past, it must be a past of a certain sort" (2013, p. 189). Haji's view is an externalist view that includes something that my own work on externalism does not – an account of moral development in actual human beings that distinguishes between "two stages of moral agency in agents like us who mature psychologically over time, roughly the initial childhood stage, and the post-childhood free agent or evolved stage" (p. 193; see Haji and Cuypers 2008). Regarding little agents, Haji is way more ambitious than I, but we are both externalists about direct moral responsibility for actions.

4. Wrapping Things up

Some readers of my previous work on moral responsibility may understandably wish that I had offered an analysis of moral responsibility for actions, come down on one side or the other of such questions as whether or not Athena is a possible being and whether or not a minoteling can be morally responsible for its first action, and presented a theory about how children come to be morally responsible agents. (Indeed, I too think it would be great if someone had settled all – or any of – these issues.) Instead, among other things, I offered what I take to be powerful counterexamples to internalist views about direct moral responsibility; rebutted various arguments for the thesis that compatibilism about moral responsibility depends on the truth of internalism; argued that compatibilist readers who share my intuitions about my (alleged) counterexamples have no good reason to worry that, if they agree with my verdicts about these cases, they will be putting their compatibilism at risk; and argued that incompatibilist believers in moral responsibility also should be externalists (Mele 2019). My aim was to eliminate internalism about direct moral responsibility from contention, thereby clearing ground for various externalist theories. If my (alleged) counterexamples are successful, internalism about direct moral responsibility is false; and as yet I see no good reason to believe that there is a plausible positive historical requirement for direct moral responsibility for actions that will add much to my negative requirement.

I do not use cases featuring little agents as counterexamples to internalism. The reason for that is simple. I take the most powerful counterexamples to internalism to feature agents who, as I put it in NFMR, were "morally responsible for having a long-standing system of values" that is radically at

odds with the values implanted by manipulators, and little children do not have long-standing systems of values.¹⁰

In Mele 2019, little children come up in two connections. First, I make the obvious point that if real human beings make the transition from being neonates who are not morally responsible for anything to being children who are morally responsible for some of their actions, then the following proposition is false: “(G) If an agent is not morally responsible for any part or property of the internal condition that issued in his A-ing, then he is not morally responsible for A” (p. 30). Second, I move on from G to the suggestion that the following proposition might be true and to the observation that it applies to instant agents: “(G*) If an agent has *no moral responsibility at all* for any part or property of the internal condition that issued in his A-ing, then he lacks *full-blown moral responsibility* for A” (p. 34). I do this primarily to prepare the way for a suggestion about why some readers may, like me, find it difficult to have intuitions about whether or not instant agents in typical stories about them are morally responsible for their actions. My suggestion highlighted the observation that these stories combine elements of a story about what may be the first action for which an ordinary child has any moral responsibility at all – a relatively trivial action, such as helping a playmate find his shoes – with elements of stories about extreme adult behavior, such as a gruesome murder (pp. 32–35). This combination is jarring.

The task I set for myself in Mele 2019 was to lay internalism about direct moral responsibility to rest. Of course, the falsity of internalism entails the truth of externalism, but the truth of externalism is compatible with a range of externalist theories of moral responsibility for actions. Exploring the various options in detail calls for a group effort. My hope for this paper is that it has clarified my reasons for stopping at my proposed negative historical requirement for direct moral responsibility for actions (rather than advancing a positive historical requirement) and clarified the relationship between the earliest actions for which ordinary children may be morally responsible and my proposed negative historical requirement.

References

- Cyr, T. 2020. “Manipulation and Constitutive Luck.” *Philosophical Studies* 177: 2381–94.
- Davidson, D. 1987. “Knowing One’s Own Mind.” *Proceedings and Addresses of the American Philosophical Association* 60: 441–58.
- Dennett, D. 1984. *Elbow Room*. Cambridge, MA: MIT Press.

10 As attentive readers will have noticed, this actually is an understatement of the sort of counterexample I take to be most powerful. As I mentioned, in addition to the property identified here, the counterexamples I take to be most powerful feature agents who had successfully striven to mold their (very good or very bad) system of values.

- Fischer, J., and M. Ravizza 1998. *Responsibility and Control: A Theory of Moral Responsibility*. New York: Cambridge University Press.
- Fischer, J. 2021. "Initial Design, Manipulation, and Moral Responsibility." *Criminal Law and Philosophy* 15: 255–70.
- Frankfurt, H. 1988. *The Importance of What We Care About*. Cambridge: Cambridge University Press.
- Frankfurt, H. 2002. "Reply to John Martin Fischer." In S. Buss and L. Overton, eds. *Contours of Agency*. Cambridge: MIT Press.
- Haji, I. 2013. "Historicism, Non-historicism, or a Mix." *Journal of Ethics* 17: 185–204.
- Haji, I., and S. Cuypers. 2008. *Moral Responsibility, Authenticity, and Education*. London: Routledge.
- McKenna, M. 2012. "Moral Responsibility, Manipulation Arguments, and History: Assessing the Resilience of Nonhistorical Compatibilism." *Journal of Ethics* 16: 145–74.
- McKenna, M. 2021. "On Manipulated Agents and History-Sensitive Compatibilism." *Criminal Law and Philosophy* 15: 285–98.
- Mele, A. 1995. *Autonomous Agents*. New York: Oxford University Press.
- Mele, A. 2006. *Free Will and Luck*. New York: Oxford University Press.
- Mele, A. 2019. *Manipulated Agents*. New York: Oxford University Press.
- Watson, G. 1999. "Soft Libertarianism and Hard Incompatibilism." *The Journal of Ethics* 3: 353–68.
- Zimmerman, D. 1999. "Born Yesterday: Personal Autonomy for Agents without a Past." *Midwest Studies in Philosophy* 23: 236–66.