Canadian Journal of Philosophy

ISSN: 0045-5091 (Print) 1911-0820 (Online) Journal homepage: https://www.tandfonline.com/loi/rcjp20

Towards a structural ownership condition on moral responsibility

Benjamin Matheson

To cite this article: Benjamin Matheson (2019) Towards a structural ownership condition on moral responsibility, Canadian Journal of Philosophy, 49:4, 458-480, DOI: 10.1080/00455091.2018.1480853

To link to this article: https://doi.org/10.1080/00455091.2018.1480853

CANADIAN JOURNAL OF PHILOSOPHY
2019, VOL. 49, NO. 4, 458–480
https://doi.org/10.1080/00455091.2018.1480853

Routledge
Taylor & Francis Group

Towards a structural ownership condition on moral responsibility

Benjamin Matheson

Department of Philosophy, Stockholm University, Stockholm, Sweden

ABSTRACT

In this paper, I propose and defend a structural ownership condition on moral responsibility. According to the condition I propose, an agent owns a mental item if and only if it is part of or is partly grounded by a coherent set of psychological states. As I discuss, other theorists have proposed or alluded to conditions like psychological coherence, but each proposal is unsatisfactory in some way. My account appeals to narrative explanation to elucidate the relevant sense of psychological coherence.

ARTICLE HISTORY Received 25 September 2017; Accepted 21 May 2018

KEYWORDS Moral responsibility; non-historicism; structuralism; manipulation; coherence; narrative; ownership

1. Introduction

Alice and Bob have the psychological profiles of fully developed human adults. Among other things, they both recognise and respond to reasons, including moral reasons, and they both act from psychological states they are identified with. Suppose Alice and Bob independently perform an action A. As it stands, they seem to be just as morally responsible as one another for their respective A-ings.[1] Suppose, however, we discover that the night before Bob had the psychological states that were essential to him A-ing implanted by advanced neuroscientists. It now seems that Bob is (at least) *less* morally responsible than Alice for A-ing.[2]

Historicists have used this seeming to support the metaphysical thesis that an agent's history partly determines whether or not she is morally responsible for her current actions. We can understand historicism as positing an *ownership* condition on moral responsibility. According to the historicist, an ownership condition must have two features. First, it should tell us which mental items (e.g. psychological states, capacities, mechanisms, and so on) *belong* to the agent such that she may be morally responsible for the

CONTACT Benjamin Matheson benjamin.matheson@philosophy.su.se Department of Philosophy, Stockholm University, Stockholm, Sweden

output of those items. Second, if the agent has a past, then it should refer to the agent's past before the time of action. Two forms of historical ownership condition have been proposed. According to positive historicism, an agent must *have* a particular sort of history.[3] According to negative historicism, an agent must *lack* a particular sort of history.[4]

In this paper, I propose and defend a *structural* ownership condition. According to the condition I propose, an agent owns a mental item if and only if it is part of or is partly grounded by a coherent set of psychological states. As I discuss, other theorists have proposed or alluded to conditions like psychological coherence, but each proposal is unsatisfactory in some way. I offer what I take to be the best understanding of psychological coherence. My account appeals to *narrative explanation* to elucidate the relevant sense of psychological coherence. I will not attempt to provide a complete set of the conditions on moral responsibility. Rather, I only wish to show that such non-historical ownership conditions ought to be given more consideration and investigation.

2. Local manipulation

Non-historicism is the thesis that only non-historical properties determine whether an agent is or isn't morally responsible at the time of action. Non-historicists can be divided into two broad camps: structuralists and reasons-responsive theorists. Structuralists require that an agent has a *properly structured psychology* when she acts in order to be morally responsible for her actions.[5] Reasons-responsiveness theorists require that an agent (or her action-producing mechanisms) are *appropriately responsiveness to reasons*.[6]

An example of a structuralist view is Frankfurt's (1971) hierarchical account. On his view, an agent is morally responsible only if she acts from a properly structured will. An agent's will is properly structured if there is an alignment between her effective first-order desires (the desires that actually move her to action) and her second-order volitions (her desires about which first-order desires she wishes to be effective). If there is such alignment then, according to Frankfurt, the agent *identifies* with her effective first-order desires.

An example of a reasons-responsive view is Fischer and Ravizza (1998) account of moderate reasons-responsiveness.[7] On their view, it isn't the agent but rather her action-producing mechanisms – such as practical deliberation, unreflective habit, and so on – that must be responsive to reasons. A strongly responsive mechanism would be responsive to *all* reasons; such responsiveness is too strong because it renders weak willed actions ones an agent cannot be morally responsible for.[8] A weakly responsive mechanism would be responsive to *only one* reason; such responsiveness is too weak because it means that even agents who are responsive to bizarre or irrational reasons could be morally responsible. Fischer and

Ravizza (1998: 62–90) thus propose moderate responsiveness. This has two components: (i) *regular receptivity* to reasons, and (ii) *weak reactivity* to reasons. (i) entails that mechanisms are receptive to a stable pattern of reasons to act, which we can know if the individual can recognise that stable pattern. (ii) entails that in at least one counterfactual sequence of events the mechanism would produce a different action than the mechanism produced in the actual sequence of events. In effect, reasons-responsiveness acts as a surrogate to an alternative possibilities condition: rather than requiring that an individual be able to act otherwise in the actual sequence of events, the reasons-responsive theorist requires that the individual act otherwise in an alternative sequence of events.

Both structuralist and reasons-responsiveness conditions are only necessary conditions on being morally responsible. Another necessary condition is the epistemic condition. A plausible understanding of this condition is that an agent must have some awareness or sensitivity to the moral status of their action. For simplicity, I set aside the epistemic condition in what follows. In all the cases I discuss, we can assume that the agents involved satisfy this condition, as well as any other necessary conditions on being morally responsible. Given this, we can treat both structuralist and reasons-responsive conditions as *sufficient* for being morally responsible for the purposes of this paper.

According to historicists, all such non-historical conditions (even when supplemented with epistemic conditions) are insufficient for moral responsibility. The following sort of case is often used to support this claim:

> Local Larry: Larry has all the capacities and competences of a fully developed human adult. One night, he goes to see The Great Garibaldi, a famous stage hypnotist. Garibaldi invites Larry on stage; Larry dutifully agrees. Larry believes that Garibaldi will cure his fear of clowns. However, knowing that Larry plans to go to the bank tomorrow, Garibaldi implants a small amount of psychological states into Larry such that the next day Larry will rob that bank. Garibaldi is no ordinary stage hypnotist. He implants the psychological states such that Larry acts on a first-order desire that aligns with a second-order volition, and such that Larry acts from a moderately reasons-responsive mechanism.[9]

Is Larry morally responsible for robbing the bank? There seems to be a strong intuition in favour of him *not* being morally responsible, despite him satisfying the conditions that non-historicists claim are sufficient (in conjunction with the other necessary conditions) for being morally responsible. Historicism offers a solution: Larry isn't morally responsible because he doesn't *own* the psychological states or mechanisms that produce his action.

I agree with historicists: Larry doesn't own the relevant psychological states or mechanisms, and so he isn't morally responsible for robbing the bank. However, I believe that historicists have hastily inferred that an

ownership condition must be a historical condition. I believe that there can also be non-historical ownership conditions. I will now propose and defend a structural ownership condition.

3. Structural ownership

According to my proposed structural ownership condition, an agent owns a mental element, such as a psychological state or mechanism, if and only if her psychology is structured in the right kind of way. Specifically:

> An agent owns a mental item M *if and only if* M is part of or partly grounded by a *coherent* set of psychological states.

Consequently, an agent owns an action A if and only if A stems from a mental item M that is part of or partly grounded by a coherent set of psychological states.

To understand this ownership condition, we must understand what a 'coherent set of psychological states' – or 'psychological coherence' for short – amounts to. We might describe Frankfurt's view as requiring coherence between the agent's effective first-order desires and her second-order volitions. This *isn't* the sense of coherence I have in mind. This is coherence between levels of desire. What I am trying to capture is a general psychological coherence, one which involves all manner of psychological states (e.g. beliefs, memories, desires, values, cares, and so on).

To get a firmer idea of what I have in mind, consider a conceptually possible individual with only one desire – namely a desire to turn on radios (cf. Quinn 1993: 236). Would such an individual count as psychologically coherent? While such an individual might not be psychologically *incoherent*, they aren't psychological coherent in the sense I have mind. The reason is that I understand psychological coherence to be a positive relation – that is, a relation that holds *between* psychological states. While the 'radioman' may have some beliefs (such that he counts as an agent), he still has insufficient psychological states to even be in the running for psychological coherence or incoherence; rather, he is psychologically *noncoherent*. Psychological coherence and incoherence, in the sense I have in mind, requires more than a few psychological states.

There are versions of structuralism about *autonomous agency* that posit a general coherence between certain types of psychological state. Laura Ekstrom (1993) defends one such view. Her conception of coherence is analogous to Keith Lehrer's (1990) conception of coherence in his account of epistemic justification. On Ekstrom's view, an agent acts autonomously if and only if she acts from an authorized (i.e. her own) preference. A preference is a desire for something that is formed in the search of, and thus aims towards, the agent's subjective conception of the good. An

authorized preference is one that coheres with the agent's 'character system', and the character system consists of preferences an agent has (at a particular time). For a preference to cohere it must be consistent with *and* mutually supported by the other preferences. Since preferences aim towards the agent's conception of the good, those preferences that cohere together are those preferences that aim towards the same conception of the good. Those preferences that aim towards different conceptions of the good are unauthorized. Could such an account of coherence be adapted into a structural ownership condition on moral responsibility?

No, because it cannot accommodate moral responsibility for actions that stem from *ambivalence*.[10] It is quite plausible that I have a preference to give to charity and preference not to give to charity. I might sometimes give money to charity and sometimes not give money to charity. Intuitively, I may be morally responsible for doing either, as Shoemaker (2015: 136) makes clear. But note that these preferences are inconsistent, so I wouldn't own at least one of these preferences according to an Ekstrom-inspired sense of 'psychological coherence'. This leads to the counterintuitive result that I cannot be morally responsible for acting from ambivalence. The reason is that, on Ekstrom's view, coherence requires consistency, and it seems that psychological inconsistency alone isn't responsibility-undermining.

Others have appealed to coherence or coherence-like conditions. Smith (2005: 256) says that a morally responsible agent must have a 'coherent psychology of a certain sort, such that there are systematic rational connections between the things that happen in her psychological life and the underlying judgments and values she accepts'. Shoemaker (2003: 115) says that cares must be part of a 'nexus of cares'.[11] Unfortunately, neither author elaborates on their suggestion. Although they don't use the ownership terminology, Arpaly and Schroeder (1999) posit an 'integration' ownership condition on moral responsibility that they do elaborate upon. On their view, those beliefs and desires that are better integrated lead to actions an agent is more responsible for, and those beliefs and desires that aren't integrated lead to actions an agent isn't morally responsible for. In other words, greater integration means more ownership, and less integration means less ownership. They say that: 'beliefs and desires are well-integrated to the extent that they are deep and do not face opposed deep beliefs and desires. An action is well-integrated to the extent that it results from such beliefs and desires' (Arpaly and Schroeder 1999: 173). However, as Shoemaker (2015: 134–139) points out, their proposal is unable to deal with moral responsibility for actions that stem from ambivalence. For example, their view implies that an agent with inconsistent desires about giving to charity isn't morally responsible for either giving to charity or not

giving to charity. Hence, attempts to develop structural ownership conditions have so far come up short.

The failure of these earlier implicit attempts is instructive. I believe it suggests that the relevant sense of psychological coherence isn't a coherence between particular tokens of a type of psychological state, but rather a *general* psychological coherence – that is, one that obtains between tokens of all types of psychological states. Given this, I don't think that a fully reductive analysis of this sense of psychological coherence is possible. To understand this sense of psychological coherence, then, we must appeal to other resources.

I propose we appeal to a particular sense of *narrative explanation* to elucidate the sense of psychological coherence that I'm trying to specify. A narrative explanation is an explanation that takes the form of a story. A narrative explanation of an action should explain why an agent performed the action she did, why that makes sense with respect to who she is at the time of action, why the agent believes she became the sort of person who performs such actions, and where that action fits into her plans for the future. In other words, a narrative explanation must reference other aspects of the agent's psychology *beyond* her motivation for acting, including other aspects of her current psychology, her memories, and her plans for the future.

While my account of narrative explanation shares a lot with extant narrative accounts, it departs from other accounts in many respects. First, most accounts of narrative (e.g. Schechtman 1996; Velleman 2006; Schroer and Schroer 2014) require that the agent *herself* provides (or be disposed to provide) the narrative. But such an account (at least when adapted to become a condition on moral responsibility) will face problems arising from the fact that individuals aren't reliable narrators of themselves. Studies in social psychology suggest that we're often mistaken about what our motivations are, for example.[12] To avoid such worries, I will propose an account according to which the agent herself needn't be the narrator.

On my view, it is rather an *ideal narrator* that must be able to provide a narrative explanation of an agent's action for that action to count as stemming from psychological coherence. It might be that the agent can also tell a story similar to the ideal narrator's, but we're most reliably able to tell if the agent acts from psychological coherence if the ideal narrator can provide a narrative explanation of the agent's action. The story that the ideal narrator tells in effect *reflects* or *represents* the agent's psychological coherence. Narrative explanations will differ in terms of their intelligibility – that is, the intuitive sense in which parts of a story fit together.[13] If the ideal narrator can provide a sufficiently intelligible narrative explanation of an action, then the action stems from psychological coherence.

Second, while I agree with Velleman (2003) that emotions are an important part of the narrative explanations, my account differs from his because he holds that the emotional cadence – that is, arousal and closure of emotions – is an important, though not essential, part of narrative explanations. It might be that my sense of narrative explanation is not a full-blooded narrative explanation, then. But this is fine for my purposes. My aim is only to illuminate a sense of psychological coherence by appeal to a sense of narrative explanation. I do not claim that my conception of narrative explanation is the only one available or should be preferred to the alternatives. Henceforth when I refer to 'narrative explanation' I refer to my own sense, unless otherwise specified.

Third, unlike Schechtman (1996) and in line with Velleman (2006) and Schroer and Schroer (2014), narrative explanations do not need to include every aspect of a person's life, but need only involve smaller stories. As I will now explain, my sense of narrative explanation *only* appeals to these smaller stories. Schroer and Schroer (2014) take these smaller stories to be the main sorts of stories that individuals tell about themselves and, on their view, these are then constitutive of an individual's personal identity over time. On their view, if a person is able to tell a story about a mental state or action and they are psychologically connected (that is, causally connected and psychologically similar)[14] to the person-stage who possessed that mental state or performed that action, then they are *narratively connected* to that person-stage. When there are overlapping chains of narrative connectedness between two person-stages, those person-stages are narratively continuous with one another. While connectedness implies continuity, continuity does not imply connectedness. So, a person-stage at t_2 may not be able to provide a narrative explanation for any of the person-stage at t_1 's mental states or actions and yet those stages are all still part of the same person because there is an overlapping chain of narrative explanations connecting them. Of course, we might be able to imagine that the person-stage at t_2 can provide a life-story that involves all its earlier stages – what we might call a 'continuity narrative explanation'.[15] The sense of narrative explanation that I have in mind, however, appeals only to the individual's current psychology, and so only connects the person to her earlier stages she is psychologically connected to – what we might call a 'connectedness narrative explanation'. Again, I don't mean to rule out other possible senses of 'narrative explanation'. I am just making my sense clear in my effort to illuminate the sense of psychological coherence I have in mind. I will provide further clarifications of my sense of narrative explanation as we proceed.

Let us now turn to an example of my sense of narrative explanation. Suppose that I give money to a homeless person. The ideal narrator might say that:

> Ben gave money to the homeless person because he's the sort of person who sometimes feels bad for the needy. He feels bad every time he walks past a homeless person and doesn't give money to them (which is often), and, after having just bought himself something to eat, he happened to have some spare change in his pocket when he saw a homeless person with a dog. Ben likes dogs. He likes dogs because he had dogs whilst growing up, and so he is inclined to help those with dogs. This is why Ben gave the homeless person some money.

Two things should be borne in mind here. First, this narrative explanation is a simplification. A full narrative explanation would explicitly reference cognitive, conative, and affective states of the subject of the story, as well as referencing the subject's plans for the future. I have made this simplification in part because I'm not a professional author, and in part because these other aspects aren't necessary for my purposes in what follows. Second, the ideal narrator is a theoretical posit. We must imagine that they have access to everyone's actual psychological states, but they needn't be able to experience an agent's first-person perspective. The purpose of the ideal narrator and the appeal to narrative explanation is simply to elucidate the sense of psychological coherence I have in mind.

Let's now consider what the ideal narrator would say about Larry. Can we imagine that the ideal narrator being able to tell an intelligible story about why Larry robbed the bank? The story must reference Larry's actual psychology at the time of action, but not just his motivation for acting. I don't think the ideal narrator can provide such an intelligible story – that is, a narrative explanation – about why Larry robbed the bank. This is because Larry isn't the sort of person who likes robbing banks. If he were the sort of person who robs banks, then Garibaldi wouldn't have needed to implant the psychological states he did. He could have instead just ensured that Larry was going to go to the bank and, if it looked like Larry might not go through with it, he could have made Larry's egoistic reasons more salient to him as he was deliberating (perhaps through well placed cues in Larry's environment). But such manipulation doesn't require the implantation of psychological states; rather, it takes advantage of pre-existing aspects of the agent's psychology, and so isn't a form of local manipulation (more on this kind of case below).[16]

The ideal narrator can, at best, say that Larry robbed the bank *because he wanted to and he wanted to want to*. After all, it is true that Larry wanted to and that he wanted to want to. But while this might explain why Larry robbed the bank, it isn't a narrative explanation because it makes reference to little beyond Larry's immediate reason for acting; hence, it is insufficiently grounded in Larry's psychology to be a narrative explanation of his action. The reason why the ideal narrator cannot provide a narrative explanation for why Larry robbed the bank is that the rest of Larry's psychology isn't

conducive towards him robbing the bank. The rest of his psychological states (by hypothesis) are such that they wouldn't lead to him robbing the bank. His robbing of the bank also doesn't fit with any of his memories. And Larry has no plans for the future that follow from him robbing the bank. Hence, there are insufficient psychological ingredients for a narrative explanation. Given that the ideal narrator cannot provide a narrative explanation about why Larry robbed the bank, it follows that Larry's robbing of the bank didn't stem from psychological coherence. Hence, according to the view I am proposing, Larry doesn't own the psychological states that led to that action, and so he isn't morally responsible for that action.

But suppose that *Harry* does have the sort of psychology that disposes him to rob banks. Harry robs banks on a regular basis and plans to rob banks in future. On a particular occasion Harry won't rob a particular bank – call it 'bank X'. Garibaldi, however, uses his powers to make Harry's bank-robbing-conducive reasons more salient, which leads to Harry deciding to and then robbing bank X. Given that he's the sort of person who robs banks, it seems plausible that the ideal narrator could provide a narrative explanation for Harry's robbing of the bank, even though Garibaldi was responsible for making his bank-robbing-conducive reasons more salient. Is this a counter-example to my proposed ownership condition?

Only if it seems intuitive that Harry is *not* morally responsible. However, this seems implausible. It seems to me that Harry does, in fact, own his action and (granting he satisfies the other relevant conditions) *is* morally responsible for it. After all, he's the sort of person who robs banks on a regular basis. It seems incidental *how* his bank-robbing-conducive reasons became more salient to him prior to him acting. Those reasons could have, for example, become more salient as a result of Harry reflecting upon a motivational poster, such as one that said 'just do it' and then Harry just did 'it' (i.e. robbed the bank). In such a story, I find that Harry is morally responsible. I am therefore inclined to think Harry is morally responsible even if another agent is responsible for making his egoistic desires, values, and mechanisms more salient to him.[17]

Of course, some might worry that because Harry seems like a victim – because he has been hypnotised – that he's clearly not morally responsible. He wouldn't have robbed the bank on that occasion if it weren't for Garibaldi, after all. I agree there's a sense in which Harry is a victim. But being a victim isn't always an excuse. For instance, those who are bullied are more likely to bully others, but that doesn't seem to necessarily excuse them if they bully others. The bullying they suffered serves as an explanation, not necessarily an excuse, of their behaviour. And even if someone tempted a person who had been bullied to bully someone else, this wouldn't necessarily get this person off the hook. It would only potentially result in another individual being morally responsible for that bullying (more on this below).

There is, of course, an element of circumstantial luck here – in the bullying case and in Harry's bank robbery on the day in question. But circumstantial luck alone ought not to excuse an individual – particularly when those actions still express morally bad aspects of an individual's psychology. Being a 'victim of circumstance' only excuses (at best) when an individual is caused to act in manner they normally wouldn't; it doesn't hold water when they are the sort of person who performs such actions. It's important here that Harry robbed bank X not just because of Garibaldi's involvement. Garibaldi might give him the final push (so to speak), but many things – such as motivational posters, suggestions from friends, or childhood memories – could have pushed Harry to rob bank X.

Because Harry is a victim of Garibaldi's, it might seem that it is Garibaldi, and not Harry, who is morally responsible for the bank being robbed. But this assumes that two individuals cannot be fully morally responsible for the same action/event. We should reject this view. Joint actions show us that it's possible for two individuals to be fully morally responsible for the same act.[18] Again, just because Harry is a victim doesn't mean he isn't morally responsible for his actions. So even if we think Garibaldi is fully morally responsible, this doesn't exclude Harry from being fully morally responsible too. Regardless of how Harry came to act, he still (among other things) expresses morally dubious values and cares that form part of his coherent psychology. Regardless of what caused those values or desires to be expressed, they belong to Harry in the sense relevant to moral responsibility.

But what about *Barry*? At t_1 he is disposed to rob banks just as much as Harry is. At t_2, however, he starts to develop his character into the kind of person who doesn't rob banks. Unfortunately, at t_3 Garibaldi manipulates Barry to rob a bank in the same sort of way as he manipulates Harry.[19] It might seem intuitive that Barry is *not* morally responsible, given that he was on the path to redemption when Garibaldi manipulated him. Notice, though, that as with other cases, his lapse may have been caused by things in his environment. It's incidental that Garibaldi caused the lapse; as before, this just means that Garibaldi is *also* morally responsible for the bank being robbed. The difference with Barry, though, is that he also deserves credit for his attempts to reform. Those acts express morally good values and desires. This marks an important difference between him and Harry. So, while I think they are just as morally responsible – in this case, blameworthy – for their respective robberies, which both express equally morally dubious values, Barry is praiseworthy for more other actions than Harry is.

Setting aside Harry, what if we imagine that Larry (the non-responsible agent from the earlier example) were implanted with sufficient psychological states such that the ideal narrator could provide a narrative explanation of Larry's action of robbing the bank? Given that Larry seems non-responsible, this seems to present a problem for my view. But narrative

explanations, as noted, take the form of a story that references the subject's psychological states beyond her motivation for acting, including her memories and plans for the future. It doesn't seem like a few psychological states would be sufficient to accomplish that; rather, to be sufficiently intelligible to count as a narrative explanation in the sense I have specified, Garibaldi would have to implant Larry with *a lot* of psychological states. This changes the case from a local manipulation case to global manipulation case. I turn to this case after considering an objection.

4. Are narrative explanations essentially historical?

Since narratives refer to an agent's past, they might seem to be essentially historical. Given this, isn't the ownership condition I have offered really another historical condition? The first thing to bear in mind is that the ownership condition I have offered is *psychological coherence*. Such coherence is explicitly non-historical; it's a property of either a person-stage of an individual or an individual at a particular time. Narrative explanations by an ideal narrator are only used to *illuminate* the kind of coherence at issue. Still, it might seem that history matters insofar as a narrative explanation must reference an individual's past. However, a narrative explanation doesn't reference the individual's *actual* past, but rather her *memories* of the past (including both immediately consciously available memories and 'buried' memories). And an individual's memories needn't be veridical. There are hypothetical cases where individuals have no past and yet they plausibly act from psychological coherence – namely, so-called instant agents.[20] The ideal narrator can plausibly provide a narrative explanation of an instant agent's very first action. They will be able to do this if the instant agent is created with the necessary ingredients for psychological coherence, and this seems possible given that psychological coherence is a relation that holds between the individual's psychological states – for example, beliefs, desires, memories, values, cares, and so on. The fact that some or all of her memories might be false doesn't matter: the narrative explanation of her action will still explain her actions in terms of psychological states beyond those that moved her to action.[21]

One might worry that the ideal narrator's narrative explanation is no more reliable an indicator of psychological coherence than the first-person narrative explanation that I dismissed earlier due to worries about confabulation. It might be that the narrator is simply telling us an intelligible story but that the agent does not in fact act from psychological coherence; it only seems like she does because of the story's reliance on implanted pseudo-memories.

In response, it is first important to note that the sense of psychological coherence I have mind is not a phenomenological sense of psychological

coherence. So, the story told by the ideal narrator is not meant to represent the agent's sense (deluded or otherwise) of her own coherence. Rather, it is meant to represent a structural sense of psychological coherence that holds between psychological states (though I do not intend to dismiss a connection between these two senses of psychological coherence). Second, notice that non-implanted memories may also be partly or entirely non-veridical, but even so they seem able to support a story about why a person acts the way she does on a particular occasion. For example, the fact a person remembers her attackers having a gun even though they had no gun helps explain why the person later testifies in court that this was the case. The fact you falsely remember me saying that I dislike carrots helps to explain why you don't serve me carrots. Such non-veridical memories play the same functional role as veridical memories do in psychological coherence. Since pseudo-memories are a type of non-veridical memory, it seems fine for them to help constitute one's psychological coherence.

Further, one might worry that since narrative explanations involve appeal to emotions and that because emotions are 'essentially diachronic' (Velleman 2003: 13) that narrative explanations must also be essentially diachronic. However, the claim that emotions are essentially diachronic is suspect. It seems clear that emotions are *dispositionally* diachronic – that is, they are disposed to unfold or persist through time. But being disposed to do something and actually doing it are two different things. It is true that I wouldn't plausibly care about something (that is, have an emotionally-laden attitude towards something) at t_1 if I wasn't disposed to continue to care about it beyond t_1. But it seems implausible that I actually have to continue to care beyond t_1 to count as caring at t_1. We might suppose that the above mentioned instant agent is brought into existence for a moment and then immediately taken out of existence. I see no reason why the instant agent cannot be said to care about things or experience emotions more generally in that moment, given that she is brought into existence with the full psychological profile of a fully developed human adult. What seems important to her having cares and emotions, it seems to me, is that she *would* have continued to have them beyond that moment if her life hadn't been tragically cut short.[22] I suspect that others have confused grounds for *attributing* an emotion or care with grounds for *possessing* an emotion or care.[23]

To support and clarify the above considerations, consider the following analogy. Normally, a history is required to make a cake (one involving certain ingredients being mixed, and the resulting batter being cooked at a certain temperature for a certain amount of time), but that process can be circumvented *in principle*. We might imagine a god-like being bringing a cake instantly into existence. The same is true, I claim, with morally responsible agents: *normally* it takes years to go from a mere individual to a

morally responsible agent, but that process can also be circumvented in principle. So, the fact that being morally responsible for an action requires psychological coherence doesn't require that a morally responsible agent has a real history (though morally responsible agents will normally have real histories). All that's required is that an individual has memories (or memory-like states). If an action issues from psychological coherence, then the agent will have such psychological states. It doesn't matter *how* the agent came to be psychologically coherent; all that matters for moral responsibility, in my view, is that her actions issue from such coherence.

To repeat, the ideal narrator's narrative explanation is reflective of the agent's psychological coherence, and not of her actual history. Hence my structural ownership condition isn't a backdoor historical condition.

5. Global manipulation

5.1 *Variant 1*

But what if an individual is implanted with large amount of psychological states? Let's now consider such 'global' manipulation cases.[24] Here is most widely discussed global manipulation case:

> *Brainwashed Beth*. 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 character – or collection of values – left no place for a desire to do such a thing to take root. ⋯ 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 [a serial killer] 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. ⋯ 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. (Mele 2013: 169–170)

It seems to many that Beth isn't morally responsible for killing George. Granting that Beth has satisfied all the leading non-historical conditions,[25] it seems that those conditions are insufficient for being morally responsible.

While we might judge that Beth isn't morally responsible, there are some important details of this case that seem problematic. By hypothesis, Beth has the background psychological contents – that is, the beliefs, desires, memories, and all the rest – of an exceptionally sweet person, yet she now has the values of a serial killer.[26] These new values lead to the production of

a desire to kill George. Mele (2013: 169) claims that Beth would be 'surprised by this new desire'. Presumably, she would also be surprised by her unexplained acquisition of the values of a serial killer. But merely being 'surprised' by her murderous desires and her serial killer values seems absurd. It seems that any previously exceptionally sweet person would be more than just surprised by this. I think they would find it downright horrifying, and wouldn't, as Mele suggests, simply take it in their stride. Beth, after all, remembers being exceptionally sweet, yet she now has the values of a serial killer that has produced a desire to murder her neighbour. While it might make sense to her why she has that desire, given her new values, it won't make sense to her why she has those values. At the very least, we would expect Beth to be severely confused by this sudden change. Indeed, it's likely that she would believe that she had a severe mental illness or she would display the symptoms of someone suffering from a severe mental illness. Beth might even attribute this nefarious change in her values to the actions of some other agent. Her attribution would be similar to how a person suffering from thought insertion attributes a thought she is having to some other agent – that is to say, the person thinks that someone else is thinking thoughts in her mind. The only difference being that Beth is correct: she does have someone else's values.

Of course, Mele might just stipulate (however dubiously) that Beth doesn't have these psychological problems when she kills George. Let's grant that stipulation for the sake of argument. Is this enough to render Beth psychologically coherent? Let's consider what story the ideal narrator would say about her.

Given that Beth has the values of a serial killer, the ideal narrator *will* be able to explain her action with respect to parts of her psychology at the time of action beyond her motivation for acting. They might even be able to explain Beth's action with respect to her planned future actions, if we assume that these plans derive solely from her new system of values. But, because Beth has some of the beliefs and all of the memories of an exceptionally sweet person, the ideal narrator cannot explain her action with respect to her memories of her past actions. All she will remember are the actions of a sweet and kind person, actions which cannot explain why she now has the values of a serial killer. There will therefore be a disparity between her current psychological profile and how she came to have that psychological profile, as her memories suggest that she shouldn't be the sort of person she now is. This disparity would render the ideal narrator's attempt at a narrative explanation of Beth's killing of George unintelligible, because that story would be muddled between Beth's memories and remaining beliefs associated with being an exceptionally sweet person, having the values, beliefs, and desires of a serial killer, and having plans for future heinous actions. In short, Beth lacks the psychological

resources for the ideal narrator to be able to tell us a sufficiently intelligible story about why Beth killed George. So, Beth doesn't satisfy my ownership condition.

5.2 *Variant 2*

Suppose we change the case so that the manipulators provide Beth with pseudo-memories such that the ideal narrator is then able to provide a narrative explanation of her action of killing George (cf. Dennett 2003: 283). Is Beth morally responsible for killing George? There is now no good reason to think she isn't. Notice that it is clear that *in this variant* post-brainwashing Beth is a *numerically distinct person* to pre-brainwashing Beth, according to the psychological continuity theory.[27] According to Parfit (1984: 206), psychological continuity requires overlapping chains of strong psychological connectedness – that is, the holding of more than 50% of the connections that normally hold from day to day in normal people. Examples of such connections include an experience and a memory of that experience, and forming a value and continuing to hold that value. In this variant, the two Beths aren't just different in terms of their values and desires, but also in terms of their memories, so there doesn't seem to be enough to maintain strong psychological connectedness and, thus, psychological continuity. Indeed, Parfit (1984: 207) notes that 'there would not be continuity of character if radical and unwanted changes were produced by abnormal interference, such as direct tampering with the brain'. Hence there is a break in psychological continuity – and hence a break in personal identity – between pre- and post-brainwashing Beth. Mele (1995: 175, n.22) does try to pre-empt worries about personal identity by claiming that the brainwashing in cases like this doesn't undercut psychological continuity. But this only seems plausible in variant 1, where post-brainwashing Beth has pre-brainwashing Beth's memories. In this variant, it seems clear, assuming a psychological account of personal identity, that post-brainwashing Beth is a numerically different person to pre-brainwashing Beth.

Of course, Mele also claims that post-brainwashing Beth is the same person as pre-brainwashing Beth according to *non-psychological* accounts of personal identity. This also seems true in this variant of his case. However, if personal identity is what determines responsibility *over time*, then it seems that non-psychological accounts of personal identity are untenable. Such accounts would imply that an individual is morally responsible for some past action even though she shares *no* psychological features with her earlier self, who performed the action. For example, suppose Garry murders someone while satisfying all the relevant conditions on moral responsibility, but then has all his psychological states erased instantly, and is then given an entirely new and different set of psychological states. If a non-psychological

account of personal identity is true and personal identity is what determines moral responsibility over time, then Garry-with-the-new-psychology is morally responsible for Garry-with-the-old-psychology's action. This is preposterous. They are psychologically alike and related in absolutely no respect; they just happen to be identical, according to non-psychological accounts of personal identity. So, either non-psychological accounts must be rejected *or* personal identity doesn't ground responsibility over time. Either way, the relevant relation for moral responsibility over time is psychological in nature. Notably, Olson (1997: 58) – a prominent defender of a non-psychological account of personal identity known as 'animalism' – agrees.

Let's suppose that a non-psychological account of personal identity is true for the sake of argument. We might then say that even though pre- and post-brainwashing Beth are numerically identical, they are *practically* distinct. We implicitly make this distinction in real life all the time. We sometimes say that an individual is 'not the person they used to be' without meaning that they are literally a new entity. On the kind of view implied by Olson, practical identity (and not numerical identity) is the relation that underlies moral responsibility over time, and practical identity is a psychological relation. This relation cannot be psychological *continuity*, because this relation allows for absolutely no psychological similarity between an individual at two times. Remember: psychological continuity only requires overlapping chains of strong psychological connectedness, so psychological continuity might hold between an individual at two times even though the individual at those two times is in no way psychological connected, and it is psychological connectedness that implies psychological similarity.

We need not specify exactly what the relation of practical identity over time is to see how this reply works. We can simply say that because there's a significant psychological difference – that is, practical identity doesn't hold – between pre- and post-brainwashing Beth that we have no grounds to treat post-brainwashing Beth like pre-brainwashing Beth (and vice versa) if, say, the brainwashing was reversed after Beth killed George. The upshot is that there's no problem with holding post-brainwashing Beth is morally responsible, because this doesn't imply that pre-brainwashing Beth (the exceptionally sweet parts of Beth) would be morally responsible for post-brainwashing Beth's actions.

5.3 *Variant 3*

If my response to variant 1 were unsuccessful – that is, it's in fact plausible that post-brainwashing Bath acts from psychological coherence – then we can extend my reply to variant 2 to this variant. To avoid confusion, let's call this 'variant 3'. This variant features the following three stipulations: (i) Beth

has no pseudo-memories implanted; (ii) post-brainwashing Beth is the same person, according to the psychological continuity theory, as pre-brainwashing Beth; (iii) post-manipulation Beth acts from psychological coherence – that is, the ideal narrator can, counter to what I argued in §5.1, provide a narrative explanation of Beth murdering George. While this seems to be how Mele intends his case to be read, it doesn't strike me as the most natural way to read this case. In effect, I've given reasons to doubt whether all three stipulations can be held consistently, but I'll now show that my reply can be extended even granting these conjointly unnatural stipulations – that is, even granting that I am incorrect that post-manipulation Beth is psychologically incoherent when she kills George.

Given that my account implies that post-manipulation Beth in this variant *owns* her action of killing George (because we've accepted for the sake of argument, however dubiously, that she acts from psychological coherence) and (granting she satisfies all other necessary conditions on being morally responsible) *is* morally responsible for doing so, my account might seem unable to accommodate the strong intuition that post-manipulation Beth isn't morally responsible. Again, I must stress that we are assuming, counter to what I argued in §5.1, that post-manipulation Beth is psychological coherent. Assuming this, though, given that there is a significant (if not numerical-identity-breaking) psychological difference between pre- and post-manipulation Beth (they have radically different values), we can still say that post-manipulation Beth isn't practically identical to pre-manipulation Beth, and any reversal of the brainwashing will lead to a similar break in practical identity. Given this, the intuition that *Beth* isn't morally responsible is undercut because it, I contend, rides on the fact that Beth is 'exceptionally sweet'. Post-brainwashing Beth isn't exceptionally sweet; for her entire (perhaps short) practical existence she has been nothing other than a moral monster. Hence, this variant isn't a counterexample to my proposed ownership condition on moral responsibility either. So, even if we assume post-manipulation Beth is psychologically coherent, this case doesn't undermine my proposed view.[28]

6. Conclusion

Consider Alice and Bob again. Alice came to have the psychological states and capacities that were essential to her A-ing under her own steam (i.e. without being manipulated). Bob, on the other hand, didn't. It therefore seems that Bob is (at least) less morally responsible than Alice; indeed, many hold that agents like him are not at all morally responsible. But this is our reaction to this case without much information on the details of the manipulation at issue, and the details matter. If Bob were locally manipulated to A – that is, implanted with a small amount of psychological states – then I

think he would be *not at all* morally responsible, because he wouldn't act from psychological coherence. So, the structural ownership condition I have posited satisfies intuitions about local manipulation. If Bob were globally manipulated, then one of three things might be true: (1) Bob might be psychological incoherent; (2) post-manipulation Bob might be numerically distinct to pre-manipulation Bob; or (3) even if post-manipulation Bob isn't numerically distinct from pre-manipulation Bob, post-manipulation Bob might be *practically* distinct from pre-manipulation Bob such that we can treat post-manipulation Bob (at least when it comes to moral responsibility) *as if* he is numerically distinct from pre-manipulation Bob. The structural ownership condition I have sketched in this paper either satisfies the intuition about a globally manipulated agent or explains it away. Either way, neither local nor global manipulation cases pose a problem for my proposed structural ownership condition on moral responsibility.

The debate between historicists and non-historicists – whether they are compatibilists or libertarians – rages on. My goal in this paper has been to show that there is room for non-historical ownership condition, and I have proposed a structural ownership condition. The debate between historicism and non-historicism hasn't been settled here. Indeed, I haven't directly criticised historicism in this paper. As with historical ownership conditions, the proposed non-historical ownership condition should be taken to be a necessary condition on being morally responsible. To ascertain a complete set of conditions on moral responsibility at the time of action, we therefore need to find other necessary conditions, such as agential conditions (e.g. reasons-responsiveness, agent-causation) and epistemic conditions (e.g. moral sensitivity or moral knowledge). That is a task for another time.[29]

Notes

1. In this paper, I am solely concerned with *direct* moral responsibility. Many distinguish between direct and derivative moral responsibility to make sense of responsibility for drunken behaviour and in so-called 'character setting' cases. As McKenna (2012: 166–167) makes clear, the distinction between direct and derivative moral responsibility is one that the non-historicist (those who hold that only factors at the time of action determine whether or not a person is responsible) can accept, because the non-historicist need only hold that there are no historical conditions (that is, factors beyond the time of action that at least partially determine whether or not a person is responsible) on direct moral responsibility. This is not to say, though, that the non-historicist *must* accept this distinction. See, for instance, Khoury (2012).
2. This case is adapted from Mele (1995: 145).
3. Positive historicists include: Fischer and Ravizza (1998) and McKenna (2016).
4. Negative historicists include: Mele (1995) and Haji and Cuypers (2007).
5. Structuralists include: Frankfurt (1971), Watson (1975), Arpaly and Schroeder (1999), Shoemaker (2003), Smith (2005), Talbert (2009), and Sripada (2016).

6. Reasons-responsive theorists include: Wolf (1987), Nelkin (2011), and McKenna (2013).
7. While Fischer and Ravizza (1998) overall view is a historical one, their historical condition is separate from their reasons-responsiveness condition, which is non-historical. On their view, an agent is morally responsible if she acts from her own moderately reasons-responsive mechanism. It's their *ownership* condition that's historical. I work with their account because it is both the most widely discussed and it accommodates an ownership condition most easily.
8. Structuralist views are often thought to struggle to accommodate moral responsibility for weak willed actions. Reasons responsive theories thus seem to have an edge over structuralist theories. However, more recent and more developed structuralist views – such as Sripada's (2016) – seem able to accommodate moral responsibility for weak willed actions. Sripada also claims his view can accommodate responsibility for out of character actions. Elsewhere, however, I argue that the 'out of character' objection fails. See Hartman and Matheson (Forthcoming).
9. Cf. Locke (1975: 104–106).
10. Given that there can be incoherent preferences, it is possible that an individual can have *differing* conceptions of the good; consequently, the agent might have *competing* character systems – i.e. independently coherent sets of preferences. While allowing multiple conceptions of the good might permit Ekstrom to accommodate the possibility of responsibility for actions that stem from ambivalence, the possibility of multiple character systems is a strange and implausible result. She would only create bigger problems for herself by attempting to avoid the initial problem in this way.
11. When an agent cares about something she is (among other things) emotionally invested in that something: she will experience positive emotions when that thing does well and negative emotions when that thing does badly; caring will also generate certain motivations. For a more developed account of cares, see Sripada (2016).
12. See Doris (2015) for an overview.
13. Schechtman (1996: 114–119) proposes an articulation or intelligibility constraint on narratives, which also applies to narrative explanations. While I think Schechtman provides a good (though far from perfect) account of intelligibility, I think that we have an intuitive understanding of what an intelligible narrative explanation amounts to that's sufficient for my current purposes. That is, we all have some sense of when stories make sense and when they don't. While Velleman (1989) talks about the intelligibility of actions, note that my concern is with the intelligibility of narrative explanations of actions.
14. I elaborate on the notion of psychological connectedness in §5.2.
15. Note while Schechtman (1996) only discusses one sort of narrative, Schechtman (2007) appeals to a similar, if not identical, distinction in response to Strawson (2004).
16. Cf. Shabo's (2010: 376) Ego Button case.
17. For more on this point, see Matheson (2016: 1979–1980).
18. See Frankfurt (1988: 54).
19. Thanks to an anonymous reviewer for pressing me to consider this case.
20. Following Davidson (1987), we might think that since instant agents lack histories they will lack thoughts and consequently not be morally responsible on that basis. Zimmerman (1999), however, has argued that such worries,

which stem from an adherence to externalism about mental content, can be circumvented without rejecting externalism about mental content. We need only suppose that an individual with a past has created the instant agent. That individual can therefore 'pass on' (so to speak) the content of their thoughts. See also McKenna's (2016) 'Suzie Instant' case.

21. Suppose the world was created five minutes ago, and everyone came into existence with (false) memories of their respective childhoods. The ideal narrator would still be able to provide narrative explanations for our actions. The fact our memories would be false doesn't mean they don't contribute to psychological coherence. See Zimmerman (1999) for more discussion of this sceptical scenario and its relation to the historicism/non-historicism debate.
22. While it is beyond the scope of this paper to discuss the details, I believe that this is also the case for grief, contra Goldie (2012) who believes that grief can only be understood a temporally extended way.
23. Thanks to an anonymous reviewer for pressing these worries.
24. This section draws from and substantially develops an argument I sketch in Matheson (2014: 329–33).
25. This is not actually clear in Mele's version of the case. In particular, he specifies that Beth cannot do otherwise because of the values she now has. I have removed this stipulation from the case because otherwise there is straightforward non-historicist response: we posit that moral responsibility requires the ability to do otherwise, and I take a reasons-responsive condition to be a compatibilist ability to do otherwise condition.
26. Of course, Beth's background psychological contents don't include those beliefs and desires that are constitutive of her values, but it certainly contains other standing beliefs and desires.
27. If Beth were implanted with a single pseudo-memory that explained in some way her overnight radical change of character – perhaps to make her think she has undergone a religious conversion – then she arguably would satisfy my ownership condition whilst being numerically the same person. This is because the ideal narrator would have a detail such that he could 'bracket off' her memories of being a good person, and so could provide an intelligible narrative explanation of why she killed George, without her having a radical enough of a change to break psychological continuity. This variant of Brainwashed Beth would be analogous to Paul's conversion on the road to Damascus. While Paul was previously a bad person, his apparent experience with God provides an explanation for why the ideal narrator can set aside his earlier memories of his character. It seems to me that Paul is morally responsible for his subsequent actions, and so it seems to be no problem to accept that this (post-manipulation) Beth *is* morally responsible in this case too (cf. also Arpaly 2003: 126–129 and Mele 2006: 179–184). Thanks to an anonymous reviewer for pressing me to consider this.
28. For specific accounts of practical identity – that is, accounts of responsibility over time – see Strawson (2011), Shoemaker (2012), Khoury (2013), Matheson (2014), and Khoury and Matheson (forthcoming).
29. Thanks to Helen Beebee and several anonymous reviewers.

Disclosure statement

No potential conflict of interest was reported by the author.

Funding

This work was supported by the Knut och Alice Wallenbergs Stiftelse (SE) (Project number: 1520110).

Notes on contributor

Benjamin Matheson received his PhD in Philosophy from the University of Manchester in 2014. He is currently a postdoctoral fellow in practical philosophy at Stockholm University with the Stockholm Centre for the Ethics of War Peace. He was previously a postdoctoral fellow in practical philosophy at the University of Gothenburg with the Gothenburg Responsibility Project. He has also been a visiting fellow at Tilburg University with the Tilburg Center for Logic, Ethics, and Philosophy of Science. He has research interests in ethics, metaphysics, moral psychology, social philosophy, and the philosophy of religion.

ORCID

Benjamin Matheson http://orcid.org/0000-0001-5047-5803

References

Arpaly, N. 2003. *Unprincipled Virtue: An Inquiry Into Moral Agency*. Oxford: Oxford University Press.

Arpaly, N., and T. Schroeder. 1999. "Praise, Blame and the Whole Self." *Philosophical Studies* 93 (2): 161–188. doi:10.1023/A:1004222928272

Davidson, D. 1987. "Knowing One's Own Mind." *Proceedings and Addresses of the American Philosophical Association* 60 (3): 441–458. doi:10.2307/3131782

Dennett, D. 2003. *Freedom Evolves*. New York: Viking Press.

Doris, J. 2015. *Talking to Our Selves: Reflection, Ignorance, and Agency*. Oxford: Oxford University Press.

Ekstrom, L. 1993. "A Coherence Theory of Autonomy." *Philosophy and Phenomenological Research* 53 (3): 599–616. doi:10.2307/2108082

Fischer, J. M., and M. Ravizza. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.

Frankfurt, H. 1971. "Freedom of the Will and the Concept of a Person." *Journal of Philosophy* 68 (1): 5–20. doi:10.2307/2024717

Frankfurt, H. 1988. *The Importance of What We Care About: Philosophical Essays*. Cambridge: Cambridge University Press.

Goldie, P. 2012. *The Mess Inside: Narrative, Emotion, and the Mind*. Oxford: Oxford University Press.

Haji, I., and S. Cuypers. 2007. "Magical Agents, Global Induction, and the Internalism/Externalism Debate." *Australasian Journal of Philosophy* 85 (3): 343–347. doi:10.1080/00048400701571602

Hartman, R., and B. Matheson. Forthcoming. *Moral Responsibility for Acting Out of Character*. Unpublished manuscript.

Khoury, A. 2012. "Responsibility, Tracing, and Consequences." *Canadian Journal of Philosophy* 42: 187–207. doi:10.1080/00455091.2012.10716774

Khoury, A. 2013. "Synchronic and Diachronic Responsibility." *Philosophical Studies* 165 (3): 735–752. doi:10.1007/s11098-012-9976-6
Khoury, A., and B. Matheson. forthcoming. "Is Blameworthiness Forever?" *Journal of the American Philosophical Association*.
Lehrer, K. 1990. *Theory of Knowledge*. Abingdon: Routledge.
Locke, D. (1975) "Three Concepts of Free Action." *Proceedings of the Aristotelian Society, Supplementary Volumes* 49: 95–125.
Matheson, B. 2014. "Compatibilism and Personal Identity." *Philosophical Studies* 170 (2): 317–334. doi:10.1007/s11098-013-0220-9
Matheson, B. 2016. "In Defence of the Four-Case Argument." *Philosophical Studies* 173 (7): 1963–1982. doi:10.1007/s11098-015-0587-x
McKenna, M. 2012. "Moral Responsibility, Manipulation Arguments, and History: Assessing the Resilience of Nonhistorical Compatibilism." *Journal of Ethics* 16 (2): 145–174. doi:10.1007/s10892-012-9125-7
McKenna, M. 2013. "'Reasons-Responsiveness, Agents, and Mechanisms'. In Shoemaker, D. (ED.)." *Oxford Studies in Agency and Responsibility* 1: 151–183.
McKenna, M. 2016. "A Modest Historical Theory of Moral Responsibility." *Journal of Ethics* 20 (1–3): 83–105. doi:10.1007/s10892-016-9227-8
Mele, A. 1995. *Autonomous Agents: From Self-Control to Autonomy*. Oxford: Oxford University Press.
Mele, A. 2006. *Free Will and Luck*. Oxford: Oxford University Press.
Mele, A. 2013. "Manipulation, Moral Responsibility, and Bullet Biting." *Journal of Ethics* 17 (3): 167–184. doi:10.1007/s10892-013-9147-9
Nelkin, D. 2011. *Making Sense of Freedom and Responsibility*. Oxford: Oxford University Press.
Olson, E. 1997. *The Human Animal*. Oxford: Oxford University Press.
Parfit, D. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
Quinn, W. 1993. *Morality and Action*. Cambridge: Cambridge University Press.
Schechtman, M. 1996. *The Constitution of Selves*. Ithaca: Cornell University Press.
Schechtman, M. 2007. "Stories, Lives, and Basic Survival: A Refinement and Defense of the Narrative View." *Royal Institute of Philosophy Supplement* 60: 155–178. doi:10.1017/S1358246107000082
Schroer, J. W., and R. Schroer. 2014. "Getting the Story Right: A Reduction Narrative Account of Personal Identity." *Philosophical Studies* 171: 445–469. doi:10.1007/s11098-014-0278-z
Shabo, S. 2010. "Uncompromising Source Incompatibilism." *Philosophy and Phenomenological Research* 80 (2): 349–383. doi:10.1111/phpr.2010.80.issue-2
Shoemaker, D. 2003. "Caring, Identification, and Agency." *Ethics* 114 (1): 88–118. doi:10.1086/376718
Shoemaker, D. 2012. "Responsibility Without Identity." *Harvard Revieiw of Philosophy* 18 (1): 109–132. doi:10.5840/harvardreview20121816
Shoemaker, D. 2015. "Ecumenical Attributability." In *The Nature of Moral Responsibility*, edited by A. Smith, M. McKenna, and C. Randolf. Oxford: Oxford University Press.
Smith, A. 2005. "Responsibility for Attitudes: Activity and Passivity in Mental Life." *Ethics* 115 (2): 236–271. doi:10.1086/426957
Sripada, C. 2016. "Self-Expression: A Deep Self Theory of Moral Responsibility." *Philosophical Studies* 173 (5): 1203–1232. doi:10.1007/s11098-015-0527-9
Strawson, G. 2004. "Against Narrativity." *Ratio* 16: 428–452. doi:10.1111/j.1467-9329.2004.00264.x

Strawson, G. 2011. *Locke on Personal Identity: Consciousness and Concernment*. Princeton: Princeton University Press.
Talbert, M. 2009. "Implanted Desires, Self-Formation, and Blame." *Journal of Ethics & Social Philosophy* 3 (2): 1–18. doi:10.26556/jesp.v3i2.33
Velleman, D. 1989. *Practical Reflection*. Princeton: Princeton University Press.
Velleman, D. 2003. "Narrative Explanation." *Philosophical Review* 112 (1): 1–25. doi:10.1215/00318108-112-1-1
Velleman, D. 2006. "The Self as Narrator." In *Self to Self: Selected Essays*, edited by D. Velleman. Cambridge: Cambridge University Press.
Watson, G. 1975. "Free Agency." *Journal of Philosophy* 72: 205–220. doi:10.2307/2024703
Wolf, S. 1987. "Sanity and the Metaphysics of Responsibility." In *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*, edited by F. D. Schoeman. Cambridge: Cambridge University Press.
Zimmerman, D. 1999. "Born Yesterday: Personal Autonomy for Agents without a Past." *Midwest Studies in Philosophy* 23 (1): 236–266. doi:10.1111/misp.1999.23.issue-1