Manipulation Arguments against Compatibilism

Derk Pereboom and Michael McKenna

The Oxford Handbook of Moral Responsibility, Dana Kay Nelkin and Derk Pereboom, eds., New York: Oxford University Press, forthcoming 2021.
Penultimate Draft

Abstract: This article surveys the debate focused on manipulation arguments against the compatibility of moral responsibility in the basic desert sense and the naturalistic determination of an action by factors beyond the agent's control. Manipulation arguments draw an analogy between such causal determination and intentional deterministic manipulation by other agents, claiming that because the intentional determination precludes responsibility, so does causal determination. The dialectical structure of these arguments is analyzed, and the main sorts of objections are discussed. Compatibilist responses to manipulation arguments can be divided into two categories. *Soft-line* replies do not resist the intuition that the manipulated agent is not morally responsible, and instead aim to show that an agent who is merely naturalistically determined differs from the manipulated agent in a way relevant to moral responsibility. *Hard-line* replies, by contrast, resist the intuition, essential to the incompatibilist's case, that the intentionally and deterministically manipulated agent is not morally responsible. Versions of each type of reply are critically assessed.

Keywords: moral responsibility; free will; determinism; compatibilism; incompatibilism; manipulation argument; hard-line reply, soft-line reply

Introduction

A widespread concern for our being morally responsible for our actions is the prospect that they are naturalistically causally determined, that is, rendered inevitable by virtue of the remote past in accord with the laws of nature, factors beyond our control. However, many compatibilists about naturalistic causal determination and moral responsibility are not strongly moved by this concern. Most people who enter into this controversy begin with the assumption that human agents are morally responsible in the basic desert sense when they knowingly do wrong. That is, they deserve to be blamed or punished *just* because they've knowingly done wrong. The justification for blaming or punishing thus does not invoke further reasons – most saliently, it does not appeal to consequentialist or contractualist reasons (Pereboom 2001, 2014; cf. Feinberg 1970, Scanlon 2013). Some of these are natural compatibilists, and for them the prospect that whenever we act we are causally determined by factors beyond our control would not change this assumption. Others are natural agnostics about compatibilism and incompatibilism, and for them the moral responsibility assumption would be challenged but not defeated by the prospect of causal determination. Incompatibilists contend that these reactions fail to face up to the implications of the causal determination of our actions by factors beyond our control, and they aspire to an effective strategy for promoting this perspective.

One prominent strategy for supporting the incompatibilist's position draws on an analogy between naturalistic causal determination and intentional causal determination by other agents. Manipulation arguments for the incompatibility of naturalistic causal determination and moral responsibility begin with an imaginary case in which an agent is intentionally deterministically manipulated by another agent to perform an action. The story specifies that the conditions compatibilists would take to be sufficient for moral responsibility are satisfied. The case is meant to elicit the intuition that, due to the nature of the manipulation, the agent is not morally responsible for what she does. It is then argued that this manipulated agent differs in no responsibility-relevant respect from agents who are naturalistically determined so to act. And because there is no responsibility-relevant difference between the intentionally manipulated agents and those who are naturalistically determined, the naturalistically determined agents aren't morally responsible for the action either. The conclusions drawn include: any normal, non-manipulated agent who is naturalistically determined to perform an action is not morally responsible for it even if she satisfies all the prominent compatibilist conditions on moral responsibility; the compatibilist conditions are insufficient for moral responsibility; and naturalistic causal determination in general is incompatible with moral responsibility.[1]

Varieties of manipulation argument

An early formulation of a manipulation argument is Richard Taylor's (1974). His argument targets a compatibilist about determinism and free action who follows Thomas Hobbes (1654), David Hume (1739/1978, 1748/2000), and A. J. Ayer (1954) in characterizing acting freely as acting without impediment, that is, acting as one desires to act. Taylor's formulation in terms of free action warrants a comment on the relation between (1) acting freely and (2) moral responsibility for acting. (1) and (2) can be understood so that it would be open whether acting freely is a necessary condition on moral responsibility, or else so that this is not open. Those who believe that there are successful examples of the sort Frankfurt (1969) devised maintain that one notion of acting freely, acting with the ability to do otherwise, is not necessary for moral responsibility. On a contrasting widespread current convention, acting freely or acting with free will is conceived as acting with whatever control is necessary for moral responsibility. Your co-authors have adopted this convention, for example in defining compatibilism and incompatibilism and in describing the target of manipulation arguments (e.g., McKenna and Pereboom 2016). The target of Taylor's argument can be understood in this way. But in this article we dispense with the reference to free action and free will, and more simply characterize the target of manipulation arguments as affirming the compatibility of moral responsibility (understood in the basic desert sense) and causal determination in general.

Here is Taylor's (1974) argument:

[1] This article covers much of the same ground as Chapter 7 of McKenna and Pereboom (2016) and is indebted to that chapter.

> But if the fulfillment of these conditions renders my behavior free—that is to say, if my behavior satisfies the conditions of free action set forth in the theory of soft determinism—then my behavior will be no less free if we assume further conditions that are perfectly consistent with those already satisfied.
>
> We may suppose further, accordingly, that while my behavior is entirely in accordance with my own volitions, and thus "free" in terms of the conception of freedom we are examining, my volitions themselves are caused. To make this graphic, we can suppose that an ingenious physiologist can induce in me any volition he pleases by pushing various buttons on an instrument to which, let us suppose, I am attached by numerous wires. All of the volitions I have in that situation are, accordingly, precisely the ones he gives me. By pushing one button, he evokes in me the volition to raise my hand; and my hand, being unimpeded, rises in response to that volition. By pushing another, he induces the volition in me to kick, and my foot, being unimpeded, kicks in response to that volition. We can even suppose that the physiologist puts a rifle in my hands, aims it at a passerby, and then, by pushing the proper button, evokes in me the volition to squeeze my finger against the trigger, whereupon the passerby falls dead of a bullet wound.
>
> This is the description of a man who is acting in accord with his inner volitions, a man whose body is unimpeded and unconstrained in its motions, these motions being the effects of his inner states. It is hardly the description of a free and responsible agent. It is the perfect description of a puppet. (Taylor 1974: 45)

The compatibilism Taylor was aiming to discredit falls short of the more robust compatibilisms more recently defended, but his strategy is suggestive of how these more recent proposals can be contested. In more contemporary manipulation arguments (e.g., Pereboom 1995, 2001; Mele 2006), an agent performs an immoral action that conforms to all of the prominent compatibilist conditions historically advocated. It's specified that the agent satisfies those proposed by Hobbes (1654) and Hume (1739/1978, 1748/2000): her action is unimpeded, that is, she acts as she wants to act, so that if she had wanted to act otherwise she would have; while the desire that motivates her to act is nevertheless not irresistible for her, and thus she does not act from internal compulsion (Ayer 1954). The agent satisfies the compatibilist condition proposed by Harry Frankfurt (1971): her effective desire (i.e., her will) to act conforms appropriately to her second-order desires for which effective desires she will have. She also conforms to the reasons-responsiveness condition advocated by John Fischer and Mark Ravizza (1998) (cf., Fischer 1982, 1994: the agent's desires can be modified by, and some of them arise from, her rational consideration of the reasons she has, and if she believed that the bad consequences for herself that would result her action would be much more severe than she actually thinks them likely to be, she would have refrained from acting for that reason); rationality conditions are advocated by Wolf 1990; Haji 1998; McKenna 2000, 2012; Nelkin 2008, 2011; Sartorio 2016, among others). The agent in addition satisfies the related condition advanced by Jay Wallace (1994): she has the general ability to grasp, apply, and regulate her actions by moral reasons. For instance, when egoistic reasons that count against acting morally

are weak, she will typically regulate his behavior by moral reasons instead. She also has the capacity reflectively to revise and develop her moral character and commitment over time, and for her actions to be governed by these moral commitments, a condition on moral responsibility Al Mele (2006) and Ish Haji (1998) defend.

Treating 'X' as a variable for some way of intentionally and deterministically manipulating an agent, here is a general schema for a manipulation argument (SMA):

1. Agent S who satisfies the compatibilist conditions on moral responsibility and is manipulated in manner X to perform action A is not morally responsible for performing A. (premise, based on intuition)
2. Agent S who satisfies the compatibilist conditions on moral responsibility and is manipulated in manner X to perform action A differs in no respect relevant to moral responsibility from any normal, non-manipulated agent who satisfies the compatibilist conditions on moral responsibility and is naturalistically determined to perform action A (no-difference premise)

C. Therefore, any normal, non-manipulated agent who satisfies the compatibilist conditions on moral responsibility and is naturalistically determined to perform action A is not morally responsible for performing A.[2]

How might compatibilists respond to SMA? They standardly resist either Premise 1 or Premise 2. To reject Premise 1 is to adopt a *hard-line reply*, since it involves the uncompromising resistance to the intuition, essential to the incompatibilist's case, that the intentionally and deterministically manipulated agent is not morally responsible. Michael McKenna (2008a) defends such a hard-line reply. To reject Premise 2 is to adopt a *soft-line reply*, one that does not resist the intuition that the manipulated agent is not morally responsible, and instead aims to show that an agent who is merely naturalistically determined differs from the manipulated agent in a way relevant to moral responsibility (the terminology is due to McKenna, 2004: 216-7; 2008a: 143-4). McKenna, who champions a hard-line reply, grants that compatibilists might win battles by adopting a soft-line reply to particular instances of manipulation arguments, which involves proposing a new compatibilist condition not satisfied by the manipulated agent. As Mele (2008) puts it, "even if the [manipulation case] is a counterexample to [the sufficiency for moral responsibility of a set of compatibilist conditions], it does not follow that incompatibilism is true. Perhaps the story is not a counterexample to some superior candidate for being a set of conceptually sufficient conditions for free action and moral responsibility that is consistent with the truth of determinism and has not yet been proposed by compatibilists." But McKenna argues that a problem for this strategy is that it is open to the incompatibilist to respond with a simple adjustment to her case of manipulation, one that accommodates the newly proposed compatibilist condition. Thus, in adopting a soft-line reply, one merely

[2] Pereboom (2001; 2014) appeals to a best explanation in prosecuting his argument, and Mele (2006: 141-4, 2008) takes issue with Pereboom on this point. Pereboom replies to Mele in (Pereboom 2007 and 2014).

forestalls a revised manipulation argument for which compatibilists, at least eventually, will not have the option of a soft-line reply.

Mele (2008) distinguishes three different forms of manipulation argument. The first, which he calls a *straight manipulation argument*, has the structure of SMA above. The second form, a *manipulation argument to the best explanation*, replaces the second premise in SMA with a best-explanation premise such as, 'the best explanation that S is not morally responsible when manipulated in manner X to perform A is that S is causally determined to do so'. Pereboom's argument (1995, 2001, 2014) has this form. The third form, an *original design argument* of the sort Mele himself has developed (2006), works not from a case of an already existing agent who is manipulated, but from an agent who is originally designed to perform some act, A. Here is his (2008) version:

> [A goddess] Diana creates a zygote Z in Mary. She combines Z's atoms as she does because she wants a certain event E to occur thirty years later. From her knowledge of the state of the universe just prior to her creating Z and the laws of nature of her deterministic universe, she deduces that a zygote with precisely Z's constitution located in Mary will develop into an ideally self-controlled agent who, in thirty years, will judge, on the basis of rational deliberation, that it is best to A and will A on the basis of that judgment, thereby bringing about E. If this agent, Ernie, has any unsheddable values at the time, they play no role in motivating his A-ing. Thirty years later, Ernie is a mentally healthy, ideally self-controlled person who regularly exercises his powers of self-control and has no relevant compelled or coercively produced attitudes. Furthermore, his beliefs are conducive to informed deliberation about all matters that concern him, and he is a reliable deliberator. So he satisfies a version of my proposed compatibilist sufficient conditions for having freely A. Compare Ernie with Bernie, who also satisfies my compatibilist sufficient conditions for free action. The zygote that developed into Bernie came to be in the normal way. A major challenge for any compatibilist who claims that Ernie A-s unfreely whereas Bernie A-s freely is to explain how the difference in the causes of the two zygotes has this consequence. Why should that historical difference matter, given the properties the two agents share?

Pereboom (1995, 2001, 2014) proposes that a manipulation argument acquires more force by virtue of setting out several such cases, the first of which features the most radical sort of manipulation consistent with the proposed compatibilist conditions and with intuitive conditions on agency, each progressively more like a fourth, which the compatibilist might envision to be ordinary and realistic, in which the action is causally determined in a natural way. An additional challenge for the compatibilist is to point out a difference between any two adjacent cases that would show why the agent might be morally responsible in the later example but not in the earlier one. Pereboom argues that this can't be done: in transitions between cases there is no responsibility-relevant difference between the modes of manipulation, so that if we are to treat the first case as one in which the manipulated agent is not free and responsible, that will carry over to the next in the series, and so forth, with the

fourth and last merely being a case in which determinism is true. The agent's non-responsibility thus generalizes from the first of the manipulation examples to the ordinary case.

In his set-up (2001, 2014), in each of the four cases Plum decides to kill White for the sake of some personal advantage, and he succeeds in doing so. The four cases exhibit different ways in which Plum's murder of White might be causally determined by factors beyond his control. In a first type of counterexample (Case 1) to the prominent compatibilist conditions, neuroscientists manipulate Plum in a way that directly affects him at the neural level, but so that his mental states and actions feature the psychological regularities and counterfactual dependencies that are compatible with ordinary agency:

> **Case 1**: A team of neuroscientists has the ability to manipulate Plum's neural states at any time by radio-like technology. In this particular case, they do so by pressing a button just before he begins to reason about his situation, which they know will produce in him a neural state that realizes a strongly egoistic reasoning process, which the neuroscientists know will deterministically result in his decision to kill White. Plum would not have killed White had the neuroscientists not intervened, since his reasoning would then not have been sufficiently egoistic to produce this decision. But at the same time, Plum's effective first-order desire to kill White conforms to his second-order desires. In addition, the process of deliberation from which the decision results is reasons-responsive; in particular, this type of process would have resulted in Plum's refraining from deciding to kill White in certain situations in which his reasons were different. His reasoning is consistent with his character because it is frequently egoistic and sometimes strongly so. Still, it is not in general exclusively egoistic, because he sometimes successfully regulates his behavior by moral reasons, especially when the egoistic reasons are relatively weak. Plum is also not constrained to act as he does, for he does not act because of an irresistible desire – the neuroscientists do not induce a desire of this sort. (Pereboom 2014)

In Case 1, Plum's action satisfies the compatibilist conditions; but intuitively, he is not morally responsible for his decision. It is causally determined by what the neuroscientists do, which is beyond his control, and this makes it intuitive that he is not morally responsible for it. Consequently, it would seem that these compatibilist conditions are not sufficient for moral responsibility -- even if all are taken together.

> **Case 2**: Plum is just like an ordinary human being, except that a team of neuroscientists programmed him at the beginning of his life. The neural realization of his reasoning process and of his decision is exactly the same as it is in Case 1 (although their causal histories are different).

Although Plum satisfies all the prominent compatibilist conditions, intuitively he is not morally responsible for his decision. So Case 2 also shows that these compatibilist conditions, either individually or in conjunction, are not sufficient for moral responsibility. Moreover, it would seem unprincipled to claim that here, by contrast with Case 1, Plum is morally responsible

because the length of time between the programming and his decision is now great enough. Whether the programming occurs a few seconds before or forty years prior to the action seems irrelevant to the question of his moral responsibility. Causal determination by what the neuroscientists do, which is beyond his control, plausibly explains Plum's not being morally responsible in the first case, and I think we are forced to say that he is not morally responsible in the second case for the same reason.

> **Case 3**: Plum is an ordinary human being, except that the training practices of his community causally determined the nature of his deliberative reasoning processes so that they are frequently but not exclusively rationally egoistic (the resulting nature of his deliberative reasoning processes are exactly as they are in Cases 1 and 2.

For the compatibilist to argue successfully that Plum is morally responsible in Case 3, he must adduce a feature of these circumstances that would explain why he is morally responsible here but not in Case 2. It seems there is no such feature. Causal determination by what the controlling agents do, which is beyond Plum's control, most plausibly explains the absence of moral responsibility in Case 2, and we should conclude that he is not morally responsible in Case 3 for the same reason.

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

Given that we are constrained to deny moral responsibility in Case 3, could Plum be responsible in this ordinary deterministic situation? It appears that there are no differences between Case 3 and Case 4 that would justify the claim that Plum is not responsible in Case 3 but is in Case 4. In both of these cases Plum satisfies the compatibilist conditions on responsibility. In each the neural realization of his reasoning process and decision is the same, although the causes differ.

From this Pereboom argues that we can conclude that causal determination by other agents was not essential to what was driving the intuition of non-responsibility in the earlier cases. Because it's intuitive that Plum is not morally responsible in Case 1, and there are no differences between Cases 1 and 2, 2 and 3, and 3 and 4 that can ground his non-responsibility in the former of each pair but not in the latter, we are led to the conclusion that he is not responsible in Case 4. Pereboom conceives this an argument to the best explanation. In his view, the salient common factor that best explains the judgement that Plum is not responsible in any of the cases, a judgment that on reflection are led to make, is that in each case he is causally determined by factors beyond his control to decide as he does.

The dialectic of manipulation arguments

An important feature of the dialectic of manipulation arguments is illuminated by Patrick Todd's reply (2013) to an objection by Stephen Kearns (2012), an objection that explicitly targets Mele's original design argument, although it also applies to other versions. Kearns's objection is that if in a manipulation case it's the manipulation that is supposed to make it intuitive that the agent is not responsible, the reason for the non-responsibility verdict does not transfer to the ordinary case, since the ordinary case involves no manipulation. But if manipulation does not have the role of making it intuitive that the agent is not responsible, then it should be possible to start the argument with a deterministic case that does not feature manipulation, and this would render the manipulation examples unnecessary. In response, Todd points out that

> … the proponent of the argument contends -- and clearly must contend—that the manipulation is irrelevant as concerns what makes the agent unfree. She instead says that the manipulation can help us see that something does make the agent unfree. In other words, she first presents the scenario (say) to an agnostic, and asks whether the agnostic thinks that the agent is free (or responsible) in that scenario. And suppose the agnostic says 'no'. She then points out that whatever would make the agent unfree in that scenario would also make the agent unfree in a qualitatively identical scenario, except in which blind natural causes have taken the place of an intentional agent. (Todd 2013: 202)

In accord with Todd's assessment, here is an account of the dialectic. Incompatibilists believe that compatibilists fail adequately to face up to the implications of causal determination. The way manipulation arguments aim to remedy this putative shortcoming is by first presenting a deterministic manipulation case with the hope that it will be more successful at eliciting a non-responsibility intuition than causal determination alone does. The next step is to argue that non-responsibility is preserved even when the manipulation is subtracted, on the ground there is no responsibility-relevant difference between the deterministic case that features manipulation and one that doesn't. The salient common element is causal determination by factors beyond the agent's control, and this feature will therefore be sufficient for non-responsibility. A deterministic manipulation case, rather than being unnecessary, would thus serve to reveal that ordinary causally determined agents lack the sort of free will at issue.

A compatibilist might deny having any intuition that an agent in a manipulation case is not morally responsible. If someone has this unconflicted intuition even upon reflection, then this manipulation argument will not have the power to persuade him. But as Haji says, "if anything is clear, the literature reveals that targeted compatibilists have, generally, not taken the Four-Case Argument to be toothless…." (Haji 2009: 127). He also points out that there is no consensus among compatibilists about the conditions that delineate when a manipulated agent is responsible, and that this lack of agreement results in the manipulation argument having a certain allure (Haji 2009: 126-27). A response that commands agreement among compatibilists

hasn't emerged, which perhaps indicates that compatibilists have intuitions about manipulation cases that prove difficult to systematize in a compatibilist theory.

Todd (2011) makes the case that proponents of manipulation arguments have assumed too heavy a burden: they do not need to make it plausible that manipulated agents are not morally responsible, but only that their responsibility is mitigated. This is because compatibilists will have as difficult a time accounting for mitigated responsibility in a manipulation case as they would have explaining non-responsibility. So if the compatibilist were to agree that Plum's blameworthiness in Case 1 is mitigated, what would, on the compatibilist view, account for this? If it is intentional deterministic manipulation, then because there is no relevant difference between intentional deterministic manipulation and natural determination, the compatibilist's position would be compromised. However, both Justin Capes (2013) and Hannah Tierney (2013) have argued that compatibilism is consistent with affirming that determinism can mitigate responsibility, since it may allow for forms of moral responsibility incompatible with manipulation while preserving the claim that causal determination is compatible with the key sort of moral responsibility. Andrew Khoury's (2014) reply to Todd points out that implicit in mitigation is a concession that the manipulated agent is responsible to some degree (for reservations about Khoury's strategy, see Tierney 2014).

Three soft-line replies

Relevant difference: Causal determination by other agents

One prominent soft-line reply to the manipulation argument is that a feature of an ordinary naturalistic deterministic case (e.g. Case 4 in Pereboom's argument) that amounts to a responsibility-relevant difference from manipulation cases is that in the ordinary case the causal determination is not brought about by other agents (Lycan 1987). The key soft-line claim is that what is generating the non-responsibility intuition in manipulation cases is not causal determination per se but causal determination by other agents.

Adam Feltz (2013), as well as Dylan Murray and Tania Lombrozo (2015), have tested this suggestion in empirical studies. Feltz found judgments of diminished moral responsibility in cases of causal determination by other agents relative to naturalistic determination. But only if the manipulation by other agents was intentional and direct, as in Case 1, did his subjects, on average, fall below the midpoint between 'strongly agree' and 'strongly' disagree. On Murray and Lombrozo's interpretation, their results show that intentional control by other agents robustly generates intuitions of absence of responsibility and free will, while causal determination per se does not do so. They conclude that because causal determination per se does not robustly generate non-responsibility intuitions, drawing a parallel to manipulation by agents fails to support incompatibilism.

However, the incompatibilist in fact predicts that immediate assessments about responsibility will generally differ between Case 1 and Case 4, but maintains that at this point a second phase of the argument becomes pertinent: a request to explain the differences in intuition in a principled way. *Should* intentional direct manipulation by other agents make a

difference relative to natural causal determination in assessing moral responsibility in the basic desert sense? It would be valuable to survey respondents taking these considerations into account. Subjects might be challenged to explain differential judgments across the cases, and then tested to see whether their judgments about the individual cases change as a result.

The incompatibilist might also object that despite the reactions of the subjects, the difference between intentional manipulation by other agents and naturalistic determination might still be irrelevant to moral responsibility in the basic desert sense. One might test this hypothesis by having subjects imagine further cases that are exactly the same as Case 1 or Case 2, except that states at issue are instead produced by a spontaneously generated machine—a machine specified to have no intelligent designer (Pereboom 2001) or a force field (Mele 2006). However, it's hard to separate the idea of a sophisticated machine from intelligent designers of that machine, even if it's specified that the machine is spontaneously generated. The mechanism by which a force field manipulates may be too unclear, and it might well suggest bypassing to at least some subjects.

In response to these concerns, Gunnar Björnsson constructed a scenario where all the prominent compatibilist conditions on moral responsibility are satisfied but in which a cause that isn't an agent — an infection — slowly renders the agent in the example increasingly egoistic without bypassing or undermining his agential capacities. Björnsson predicted that if subjects were prompted to view the agent's behavior as dependent on this non-agential cause, attributions of responsibility would be undermined to roughly the same extent as in cases of an intentional manipulator. This turned out to be true: in a study involving 416 subjects, the infection undermined attributions of free will and moral responsibility to the same degree as in cases of intentional manipulation (Björnsson and Pereboom 2016). Björnsson's study suggests that incompatibilists might be able to employ the generalization strategy used in manipulation arguments without introducing intentional manipulation, and thus avoiding the experimentally-driven objections developed by Feltz, Murray and Lombrozo.

Relevant difference: No agency in Case 1

A different soft-line reply to Pereboom's argument has been proposed by a number of soft-line critics (Fischer 2004, Mele 2005, Baker 2006, Demetriou 2010) of Pereboom's argument. In his Case 1, Plum was created by a team of neuroscientists who can manipulate him directly through radio-like technology, but he is as much like an ordinary human being as is possible, given his history. The scientists locally manipulate him from moment to moment, by pushing buttons on their device. In this particular situation, by doing so they cause him to undertake an egoistic process of reasoning which deterministically leads to his killing of White (Pereboom 2001: 112-3). The concern raised by the critics is that in this first case Plum's mental states lack the causal cohesion required for him to be a genuine agent. The specific worry raised by Demetriou (2010) is that when the scientists manipulate Plum at some particular instant, they suppress the causal efficacy of his prior mental states, thereby making it the case that Plum's mental states are causally isolated from one another (Demetriou 2010: 608; cf., Matheson 2016: 1972). While Plum in Case 1 and Plum in Case 4 have the same mental states,

these states differ radically with regard to how they are causally connected. This difference yields a difference in agency and, potentially, in moral responsibility.

In response, Pereboom (2014) embellished the description of Case 1 so that the neuroscientists manipulate him only sporadically. In this particular instance they [the team of neuroscientists] do so by pressing a particular button just before he begins to reason about his situation, which they know will produce in him a neural state that realizes a strongly egoistic reasoning process, which the neuroscientist know will deterministically result in his decision to kill White (as suggested by Shabo 2010: 376). Pereboom suggests that in ordinary life, external influences, such as finding out that the home team lost, have a similar sort of effect, and in such situations agency is clearly preserved. In embellished Case 1, while agency is preserved, Plum is intuitively not morally responsible since had the neuroscientists not intervened, he would not have killed White – or at least so Pereboom contends (Pereboom 2014: 76).

McKenna criticized this embellished case on the ground that it is much easier to accept that Plum is morally responsible here than it is for the version on which he is constantly being manipulated. This stands to reason, but Pereboom believes that the embellished case is strong enough. Imagine that Plum is on trial by jury, and that it was revealed that the neuroscientists intervened as they did, and that he would not have killed White had it not been for their intervention. What jury would convict him? McKenna (2008, 2014) contends that a better version of Case 1 features regular local manipulation while retaining the causal integration of Plum's states. Plum in Case 1 so acts as a result of numerous "micro" interventions that "steer" Plum in certain directions rather than others while leaving Plum's identity and agency intact.[3] Benjamin Matheson (2016) develops a detailed account of how this might work, specifically with Demetriou's critique in mind. A key feature of his account is this: at times we are motivated to perform an action, but not sufficiently strongly enough to do so without someone else's badgering. I may want to reorganize the basement, but I won't do it unless my wife presses me to do it. In this example, my earlier mental states are not suppressed, but rather their motivational efficacy is enhanced, and my agency is clearly retained. So similarly, the neuroscientists, rather than suppressing the efficacy of all of Plum's relevant mental states, strengthen some of them and diminish the efficacy of others, and he retains his agency.

These proposals for defending Pereboom's manipulation Case 1 – Pereboom's, McKenna's, and Matheson's – lend support to McKenna's contention about soft-line replies: they are unstable because a slight modification by her opponent to the disputed example can thwart the reply. Such a modification argument simply builds in the compatibilist condition that in an earlier iteration the soft-line critic found to be lacking. This suggests that for compatibilists to effectively respond to well-executed versions of manipulation arguments, they should be prepared to defend a hard-line reply. In this case, it would seem that they must be prepared to

[3] These interventions need be no more invasive than that due to changes in hormone levels or fluctuations in blood sugar.

resist Pereboom's plausible claim that in Case 1 Plum does not act freely and is not morally responsible.

Relevant difference: The manipulated agent is not the causal source of the action

Interventionist theories of causation claim that one variable is the cause of another just in case an intervention on the first would lead to a change on the second.[4] Several compatibilists, including Oisin Deery and Eddy Nahmias (2017) and Marius Usher (2020) have used this theory of causation to design responses to manipulation arguments. Here we focus on Deery and Nahmias' version. The manipulation argument they criticize features two contrasting cases devised by Al Mele (2013):

> First, imagine Danny. One evening in 1986, Danny's parents made love, hoping to conceive a child. They got lucky. A zygote was formed (at time t1), and nine months later Danny was born. Thirty years later, Danny is walking down a deserted street and he finds a wallet with the owner's ID in it and $500. Danny takes himself to have good reasons for keeping the money, but also for returning the wallet. He deliberates for a while, and in the end he decides to keep the money, and he does so (at time t30). Assume that this occurs in a deterministic universe—that is, a universe in which, for each event E, the laws of nature and some set of events that occurred prior to E are such that these events cause E to occur with probability 1. If determinism is true, then some set of events prior to Danny's act of stealing the wallet at t30 are (together with the laws) such that they cause his deliberating and acting in that way, at that time, with probability. (Deery and Nahmias 2017, 1257)

Contrast this with the following variation:

> … a powerful Goddess, Diana, has the power to know what will happen in the future and to act in ways that ensure that specific events occur in the distant future. Diana has these abilities in part because she exists in a deterministic universe and is able to get enough information about events occurring in it (e.g., at t1) to deduce exactly what she needs to do at that time to ensure that a particular event occurs thirty years later. In this case, Diana assembles atoms in a specific way at t1 so as to create a zygote that develops into a child, grows up, finds a wallet thirty years later, and at t30 decides to keep the money it contains. For some reason, Diana wants to ensure that this event occurs at t30, and she possesses the power to alter events at t1 precisely so that she ensures that it does occur. As it turns out, the life of this intentionally created person (whom we will call Manny since he is Man-ipulated) follows the exact same course as the life of deterministic Danny (as described above). (Deery and Nahmias 2017, 1257)

Advocates of manipulation arguments contend that Manny is not morally responsible for stealing the money – they have the non-responsibility intuition in the manipulation case. They also stress that Manny and Danny each satisfy the compatibilist sufficient conditions for free

[4] This view has been developed and defended by Judea Pearl (2000, 2009) in computer science and James Woodward (2003, 2006) in philosophy.

will and moral responsibility, and contend that as a result there is no responsibility-relevant difference between Manny and Danny.

On an interventionist approach to causation, both Manny's and Danny's compatibilist agential structure, that is, their intentions and powers to steal conceived as compatible with causal determinism, count as actual causes of their stealing the money. Changes in the value of the variables representing these intentions and powers result in changes to the value of the variables representing their stealing the money. But Deery and Nahmias argue that there is a relevant difference between how the intentions and powers cause their actions in the two cases. While Danny's intention and powers are the causal source of his action, Manny's are not, and Manny can't be free and morally responsible with regard to an action if he, by virtue of his intention and powers, isn't its causal source (Deery and Nahmias 2017, 1267)

Deery and Nahmias contend that for a variable to represent the causal source of an event, that variable must bear the strongest causal invariance relation to that event among all the variables that are causally related to it, and,

> a causal invariance relation, R1, that obtains between two causal variables, X and Y, is stronger than another such relation, R2, obtaining between Y and another of its prior causal variables – for instance, W – iff:
> (1) holding fixed the relevant background conditions, C, R1 predicts the value of Y under a wider range of interventions on X than R2 does under interventions on W; and
> (2) R1 predicts the value of Y across a wider range of relevant changes to the values of C than R2 does. (Deery and Nahmias 2017, 1262–1263)

Danny's intention and powers are the causal source of his stealing the money because changes in the value of the variable representing Danny's intention and powers to steal issue in the same action across many changes in the background conditions, and there is greater invariance for this variable than for any competing variable. However – and this is the crucial point – Deery and Nahmias contend that there is a stronger causal invariance relation between the variable representing Diana's intention and powers to manipulate and the one representing Manny's stealing than between the variables representing Manny's intention and powers and his stealing. Diana intends for Manny to steal the money at t30 and she is able to ensure that he does so. There is only a limited variety of changes in the background circumstances that would break the causal connection between the variables representing Diana's intention and Manny stealing the money. By contrast, there are a comparatively more changes in the background circumstances that would break the link between the variables representing Manny's intention and powers and his stealing the money. Consequently, Diana's intention and powers relate more invariantly to Manny's stealing the money than do Manny's intention and powers, and for this reason Diana is the causal source of the action, and not Manny.

In response, one factor that Deery and Nahmias do not take into account is that in constructing a manipulation case, both the nature of the manipulation and the nature of the manipulated agent can be made to vary, with the result that the two can be made to match in degrees and kinds of invariance. Suppose Diana is quite powerful, but not omniscient or omnipotent, but due to her impending death, she has only one shot at manipulating Manny. She does what's specified in Deery and Nahmias description, but due to lack of information, she believes she has only a small chance of succeeding. But what she does succeeds in causally

determining Manny to steal. Let's imagine the other possibilities for manipulation she was considering would have failed. In this case she can't ensure that Manny steals, but our intuition that Manny isn't responsible is still in place. However, in this example, it's false that there are very few changes in background circumstances that could break the causal connection between Diana's intention and Manny stealing the money. Moreover, Manny's abilities can be specified so that the degrees of invariance relative his stealing the money match hers precisely. In this scenario, it's false, on the interventionist approach, that Manny is not the causal source of his stealing.

The following more vivid case might make the point even clearer. Diana, as it happens, has considerable powers to affect things from learning details about the laws of nature and the state of the world at a time. But she has no knowledge of these things at all, and for the most part she cannot gain access to such knowledge. Instead, she works for Zeus as a clerk doing mostly menial tasks. For whatever reason, Zeus has a very odd system of managing all his knowledge of the state of the world and how any possible future will unfold. He keeps every such sequence of possible ways a total history of the world will unfold recorded in a set of sheets of paper. Each such sequence is printed out (in astoundingly tiny print) on single pages. He also carries them all about in a big manila folder, maybe to show off to everyone how much he knows. One day, Zeus is walking down the hallway near her office, and one piece of paper happens to fall from his folder without his noticing it. Diana picks it up, and it is a sequence that, if actualized, will result in Manny's misdeed. Out of nothing but perverse interest, like the desire to see the fall of an arrangement of dominos, Diana decides it would be amusing to witness Manny perform this moral wrong as an upshot of this one way the world might unfold. Based on this one-off bit of knowledge that she pilfered from Zeus, she intervenes in the unfolding the world to get the desired result. Manny steals the money at t30. Here it would seem that Manny is the cause of his stealing by interventionist standards, and that the intuition that he is not responsible would be widespread.

Here is a resulting lesson for constructing manipulation arguments (Pereboom 2014): set up the cases so as to secure as close as possible a causal match between the manipulation cases and the natural deterministic case, while preserving the intuition of non-responsibility in the manipulation cases. Applying this rule, as in the cases as just described, the resulting manipulation matches natural determination better than the Mele cases as Deery and Nahmias interpret them.[5]

In fairness to Deery and Nahmias, it bears mentioning that weakening manipulator's invariance as in the above cases may also weaken intuitions of non-responsibility. Indeed, by Deery and Nahmias's standards, a compatibilist might protest that in these cases Manny remains the causal source of his wrongdoing, not Diana. So, the compatibilist might contend

[5] A different sort of objection to this strategy is raised by Hannah Tierney and David Glick (2020). They criticize Deery and Nahmias's analysis of sourcehood by drawing a distinction between two forms of causal invariance that can come into conflict on their account. They conclude that any attempt to resolve this conflict will either result in counterintuitive attributions of moral responsibility or will undermine Deery and Nahmias's response to manipulation arguments.

that Manny's moral responsibility is not compromised. However, a proponent of the manipulation argument will disagree. She can grant that the intuition of non-responsibility in cases of this sort might be weaker than it is in a case wherein the manipulator is strongly invariant and so is the causal source of the manipulated agent's targeted wrongdoing. Nevertheless, she can argue, it is strong enough to elicit a judgment of non-responsibility in an audience of neutral inquirers. Still, in our estimation, it counts in favor of Deery and Nahmias's defense of compatibilism that they have put pressure on the incompatibilist to revise her argument in a way that results in a somewhat less compelling example of manipulation as it figures in the first premise of SMA.

A hard-line reply

What if the compatibilist claims that it's intuitive that Plum is morally responsible in Case 4, and because there are no responsibility-relevant differences between adjacent cases, the verdict of responsibility carries backward through to Case 1? Then the compatibilist and the incompatibilist would be at a standoff, with the manipulation argument yielding progress for the incompatibilist. However, for the compatibilist to claim that it's intuitive that Plum is in fact morally responsible in Case 4 would be overly ambitious, since that's what's fundamentally at issue in the debate. McKenna's (2008) hard-line reply proposes instead that an *agnostic* stance about Case 4 is initially reasonable, and that this agnostic stance transfers to Case 1 unimpeded due to the absence of responsibility-relevant differences between adjacent cases. Credible doubt is accordingly cast on the claim that Plum is not morally responsible in Case 1. In this way, McKenna addresses the multiple-case argument with a less ambitious but more compelling hard-line strategy.

In developing this strategy, McKenna (2008a; 2014) first invites us to consider pertinent agential and moral properties of Plum as displayed in Case 4, in which Plum's action is naturalistically determined. [6] There Plum is just as much like an ordinary moral agent as any one of us. He is capable of feeling remorse, or instead ambivalence about killing White, or even pride in that accomplishment. We may in addition imagine that Plum has a history of moral development from childhood into adulthood like that of any normal person. We can suppose that Plum features the emotional complexity that Strawsonians highlight – he is capable of moral resentment and gratitude, indignation and approval, guilt and pride, and he is sensitive to the significance of these emotions in his relations with others. Plum is, in consequence, a full member of the moral community, one whose standing enables the personal relationships characteristic of fulfilled human lives. At this point McKenna argues that if all of this is true of Plum in Case 4, it can equally as well be true of Plum in Case 1. In his view, the point of attending to these details is to help elicit intuitions that are friendly to compatibilists—to elicit the sense that Plum is in Case 1 is a complex agent just like any normal person we might encounter. Sofia Jeppsson (2019) reinforces this part of the strategy by arguing that in assessing manipulated Plum's moral responsibility, taking on his agential and moral perspective is the

[6] This strategy, as recounted here, is most completely spelled out in McKenna (2014).

natural and the morally obligatory one for ordinary, everyday interaction, and that an incompatibilist who wants to claim that we ought to make moral responsibility judgments from a detached causal perspective therefore owes us an argument for this claim.

Next, McKenna has us consider the initial claims the compatibilist is entitled to make regarding Plum in Case 4. She is not entitled to insist that the incompatibilist grant that here Plum *is* morally responsible. Rather, she can legitimately expect that a neutral inquirer, who is undecided whether determinism rules out moral responsibility, initially allow that it's open that Plum is morally responsible in Case 4 (and that it's open that he isn't). McKenna then points out that incompatibilists may not assume at the outset that Plum in Case 4 is not morally responsible, nor should they to expect that the open-minded neutral inquirer's intuition about Plum in Case 4 is that he is not morally responsible. Again, that is what is in dispute, and assuming this at the outset would be question-begging. Thus, the stance towards Case 4 that an incompatibilist should take is the same as the one that a compatibilist should take. Both must initially allow that in Case 4 it is an open whether Plum is morally responsible for killing White (and that it's open that he's not).

With this in mind, McKenna attempts to turn Pereboom's strategy against him. Pereboom[7], assuming an intuition of non-responsibility in Case 1, marches through to his Case 4, drawing an incompatibilist conclusion. He grants that Pereboom does gain *some* intuitive advantage in favor of incompatibilism by beginning with the non-responsibility intuition in Case 1 and having this verdict transmit through to Case 4. That has to be weighed on the scales when evaluating the overall force of the argument. But McKenna proposes that the compatibilist now move in the opposite direction, and that the result also be weighed on the scales. It's open that Plum in Case 4 is morally responsible; there are no responsibility-relevant differences between adjacent cases; we can thus conclude that it's open that Plum is morally responsible in Case 1.

Finally, McKenna aims to mitigate the potentially jarring effect of claiming that it's open that agents in manipulation cases are morally responsible by drawing our attention to actual, fairly common events in human life similar in key respects to manipulation cases. In such actual cases, McKenna contends, our intuitions do not favor an incompatibilist diagnosis. Nomy Arpaly suggests that there are many kinds of circumstance in which agents undergo dramatic changes that affect how they act, such as a religious conversion, the birth of a child, a traumatic accident, or the death of a loved one (Arpaly, 2006: 112-3). If such changes leave the agent sane and rational, allowing satisfaction of the compatibilist conditions on moral responsibility, few would be inclined to judge that they undermine moral responsibility. McKenna suggests that reflection on the similarity of such actual circumstances to manipulation adds credence to his compatibilist strategy.

Pereboom (2008, 2014) responds that from the point of view of a neutral party, one not initially committed to compatibilism or to incompatibilism, it is initially epistemically rational not to believe that the agent Case 4 is morally responsible, and not to believe that he isn't, but

[7] See also Sekatskaya (2019).

to be open to clarifying considerations that would make one or the other of these beliefs rational. This *neutral inquiring stance* differs from the stance of the *confirmed agnostic*, who maintains that it is not clearly the case that the ordinary determined agent is morally responsible, and that it is not clearly the case that he is isn't, while it's rational to consider enquiry into the issue closed, and thus not open to further clarifying considerations. If the confirmed agnostic's stance were rational, McKenna's conclusion could be generated. But Pereboom argues that on the neutral inquiring stance, a manipulation case can function as a clarifying consideration that makes rational the belief that the ordinary determined agent is not morally responsible. In marching backward through the cases from Case 4, the intuition that it's open that Plum is morally responsible might well not survive the clarifying considerations that the manipulation cases provide. As Pereboom sees it, intuitions in Cases 1-3 are friendly to incompatibilism, and these cases provide the right sort of analogy to Case 4 to help clarify the rationality of claim it's open that Plum is morally responsible. This is so, even granting that attention to the agential and moral properties of Plum in Case 1, carrying over, as McKenna points recommends, from Case 4, do lend some weight to a compatibilist diagnosis.[8]

The dispute between Pereboom and McKenna—that is, the dispute between us, your coauthors—appears to hinge primarily upon these two interrelated questions: in light of cases like Case 1 and Case 2, has the neutral inquiring audience been given sufficient reason to move from open-minded agnosticism towards an incompatibilist verdict? Or is McKenna's hard-line strategy sufficient to establish that it remains rational to maintain that it is an open whether determined agents are morally responsible?

Manipulation and praiseworthiness

Manipulation arguments typically focus on immoral actions, on agents who would be blameworthy rather than praiseworthy. Dana Nelkin (2011), who together with Susan Wolf (1990) endorses an asymmetry thesis, that access to alternative possibilities is required for blameworthiness but not for praiseworthiness, asks us to consider a manipulated agent, like Plum in Case 1, who does the right thing for the right reasons:

> The neuroscientists have created Plum so that he fully understands and recognizes good reasons for acting. He is moved by people in distress and desires to help them so as to relieve their suffering. Suppose that one day he finds himself in a situation in which he can help a child only at great risk to himself. He thinks about the relevant considerations and decides that all things considered, helping is the right thing to do and resolves to do it. (Nelkin 2011, 56-7)

Nelkin reports that for her the intuitive force of this example seems to me to count in favor of the appropriateness of praise. Pereboom (2014, 102-3) agrees, but suggests that praise can at times be an expression of approbation or delight about the goodness of the actions or accomplishments of another, which would be appropriate even if it is not basically deserved.

[8] See Daniel Haas (2012) and Sekatskaya (2019) for defenses of McKenna on this point, and Pereboom (2014) for a reply to Haas.

Nevertheless, intuitions about this case would appear to provide support for the asymmetry thesis that Wolf and Nelkin defend.

Final words

One last question whether the sorts of real-life situations Arpaly and McKenna adduce should carry greater weight than the artificial manipulation cases in assessing the effect of external causes on moral responsibility. McKenna (2008a; 2014) contends that we should assign greater value to real-life cases; Pereboom (2008; 2014) believes we should assign greater value to the artificial cases. The Spinozistic concern he invokes at this point is that in ordinary life such judgments will have been shaped by ignorance of hidden deterministic causes (Spinoza 1677/1985), which are brought to the fore in manipulation cases. Manipulation cases may be artificial, but the artificiality is required to make the deterministic causation salient. Dana Nelkin concurs, and suggests that "one might argue that their unrealistic quality helps ensure that we are focused on the stipulated features, and that we aren't implicitly but unconsciously relying on background assumptions that we bring to ordinary life. In this way, the intuitions are arguably *more* reliable than the real-life ones" (Nelkin 2013: 125).

At some time in the near future remote local manipulation, as in featured in Case 1, may become actual. At a Material Research Society Conference in 2019, a Cornell team presented a paper that describes "the fabrication and characterization of wireless, implantable devices for the electrical stimulation of neural tissue." The background research focuses on a neuronal disorder in mice that leads to obesity, in which neurons active in indicating that enough food has been ingested aren't adequately stimulated. The implanted device, activated by remote control, provides a fix by causing the requisite neurons to fire.[9] When such devices are implanted in human beings, and are used to induce morally salient behavior, we'll have the opportunity to test our intuitions in actual, and not merely fictional cases.

Bibliography

Arpaly, Nomy. (2006). *Merit, Meaning, and Human Bondage*, Princeton, NJ, Princeton University Press.

Ayer, Alfred J. (1954). "Freedom and Necessity," in A. J. Ayer, *Philosophical Essays*, London: Macmillan, pp. 271-84.

Baker, Lynne R. (2006). "Moral Responsibility without Libertarianism," *Noûs* 40, pp. 307-30.

Björnsson, Gunnar, and Derk Pereboom. (2016). "Traditional and Experimental Approaches to Free Will," in *The Blackwell Companion to Experimental Philosophy*, Wesley Buckwalter and Justin Sytsma, eds., Oxford: Blackwell Publishers, pp. 142-57.

[9] https://mrsspring2019.zerista.com/event/member/557804

Capes, Justin. 2013. “Mitigating Soft Compatibilism,” *Philosophy and Phenomenological Research* 87: 640-63.

Deery, Oisín, and Eddy Nahmias (2017). “Defeating manipulation arguments: Interventionist causation and compatibilist sourcehood,” *Philosophical Studies* 174(5): 1255–1276.

Demetriou, Kristin. (2010). “The Soft-Line Solution to Pereboom’s Four-Case Argument,” *Australasian Journal of Philoso*phy 88, pp. 595-617.

Feinberg, Joel. (1970). “Justice and Personal Desert,” in *Doing and Deserving*, Princeton: Princeton University Press.

Fischer, John Martin. (1982). “Responsibility and Control,” *Journal of Philosophy* 79, pp. 24-40.

Fischer, John Martin. (1994). *The Metaphysics of Free Will*, Oxford, Blackwell Publishers.

Fischer, John Martin. (2004). “Responsibility and Manipulation,” *The Journal of Ethics* 8, pp. 145-77.

Fischer, John Martin, and Mark Ravizza. (1998). *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge: Cambridge University Press.

Frankfurt, Harry. (1969). "Alternate Possibilities and Moral Responsibility,” *Journal of Philosophy* 66, pp. 829-39.

Frankfurt, Harry. (1971). "Freedom of the Will and the Concept of a Person," *Journal of Philosophy* 68, pp. 5-20.

Haas, Daniel. (2013). “In Defense of Hard-line replies to the Multiple-case Manipulation Argument,” *Philosophical Studies* 163, pp. 797-811.

Haji, Ishtiyaque. (1998). *Moral Accountability*, New York: Oxford University Press.

Hobbes, Thomas. (1654). *Of Libertie and Necessity: A treatise, wherein all controversie concerning predestination, election, free-will, grace, merits, reprobation, &c., is fully decided and cleared, in answer to a treatise written by the Bishop of London-Derry, on the same subject,* London, printed by W. B. for F. Eaglesfield.

Hume, David. (1739/1978). *A Treatise of Human Nature*, Oxford: Oxford University Press.

Hume, David. (1748/2000). *An Enquiry Concerning Human Understanding*, Oxford: Oxford University Press.

Jeppsson, Sofia. (2019). “The Agential Perspective: A Hard-Line Reply to the Four-Case Manipulation Argument,” *Philosophical Studies*. https://doi.org/10.1007/s11098-019-01292-2

Kane, Robert. (1996). *The Significance of Free Will,* New York: Oxford University Press.

Kearns, Stephen. (2012). “Aborting the Zygote Argument.” *Philosophical Studies* 160 (3): 379-89.

Khoury, Andrew (2014). “Manipulation and Mitigation,” *Philosophical Studies* 168 (1): 283-94.

Lycan, William. G. (1987). *Consciousness*, Cambridge: MIT Press University Press.

McKenna, M. (2000). “Assessing Reasons-Responsive Compatibilism,” *International Journal of Philosophical Studies* 8, pp. 89-114.

McKenna, Michael. (2008). “A Hard-Line Reply to Pereboom’s Four-Case Argument,” *Philosophy and Phenomenological Research* 77, pp. 142-59.

McKenna, Michael. (2012). *Conversation and Responsibility*, New York: Oxford University Press.

McKenna, Michael. (2014). “Resisting the Manipulation Argument: A Hard-Liner Takes it on the Chin,” *Philosophy and Phenomenological Research* 89: 467-84.

McKenna, Michael and Derk Pereboom. (2016). *Free Will: A Contemporary Introduction*, London: Routledge, 2016.

Mele, Alfred. (2006) *Free Will and Luck*, New York: Oxford University Press.

Mele, Alfred. (2008). “Manipulation, Compatibilism, and Moral Responsibility,” *The Journal of* 12: 263–86.

Nelkin, Dana. (2008) “Responsibility and Rational Abilities: Defending an Asymmetrical View,” in *Pacific Philosophical Quarterly* 89, pp. 497-515.

Nelkin, Dana. (2011). *Making Sense of Freedom and Responsibility*, Oxford: Oxford University Press.

Nelkin, Dana. (2013). “Reply to Critics,” *Philosophical Studies* 163: 123-31.

Pearl, Judea. (2000, 2009). *Causality*. Cambridge: Cambridge University Press; first edition 2000, second edition 2009.

Pereboom, Derk. (1995). "Determinism *Al Dente*," *Noûs* 29, pp. 21-45.

Pereboom, Derk. (2007). “On Mele’s Free Will and Luck,” *Philosophical Explorations* 10: 163-72.

Pereboom, Derk. (2008). “A Hard-Line Reply to the Multiple-Case Manipulation Argument,” *Philosophy and Phenomenological Research* 77 (2008), pp. 160-70

Pereboom, Derk. (2014). *Free Will, Agency, and Meaning in Life*, Oxford: Oxford University Press.

Sartorio, Carolina. (2016). *Causation and Free Will*. New York: Oxford University Press.

Scanlon, T. M. (2013). “Giving Desert Its Due,” *Philosophical Explorations* 16, pp. 101-16.

Sekatskaya, Maria. (2019). “Double Defence Against Multiple Case Manipulation Arguments”, *Philosophia* 47(4), pp. 1283–1295.

Shabo, Seth. (2010). “Uncompromising Source Incompatibilism,” *Philosophy and Phenomenological* Research 80, pp. 349-83.

Spinoza, Benedict. (1677/1985). *Ethics*, in *The Collected Works of Spinoza*, Edwin Curley, ed. and tr., Volume 1, Princeton NJ: Princeton University Press.

Strawson, Peter F. (1962). "Freedom and Resentment," *Proceedings of the British Academy* 48, pp. 187-211.

Taylor, Richard. (1974). *Metaphysics,* 4th ed., Englewood Cliffs: Prentice-Hall.

Tierney, Hannah. (2013). “A Maneuver around the Modified Manipulation Argument,” *Philosophical Studies* 165, pp. 753-633.

Tierney, Hannah. (2014). “Taking it Head-on: How to Best Handle the Modified Manipulation Argument,” *Journal of Value Inquiry* 48: 663-75.

Tierney, Hannah, and David Glick. (2020). “Desperately Seeking Sourcehood,” *Philosophical Studies* 177 (4):953-970.

Todd, Patrick. (2011). “A New Approach to Manipulation Arguments,” *Philosophical Studies* 152, pp. 127-133.

Todd, Patrick. (2012). “Defending (a Modified Version of the) Zygote Argument,” *Philosophical Studies* 164, pp. 189-203.

Usher, Marius. (2020). “Agency, Teleological Control and Reliable Causation,” *Philosophy and Phenomenological Research* 100(2), pp. 302-24.

Wallace, R. Jay. (1994). *Responsibility and the Moral Sentiments*, Cambridge: Harvard University Press.

Wolf, Susan. (1990). *Freedom within Reason*, Oxford: Oxford University Press.

Woodward, James. (2003). *Making things happen: A theory of causal explanation*, New York: Oxford University Press.

Woodward, James. (2006). "Sensitive and Insensitive Causation," *The Philosophical Review* 115, pp. 1–50.