(Forthcoming in *Ergo: An Open Access Journal of Philosophy*) Direct Manipulation Undermines *Intentional* Agency (Not Just *Free* Agency)

ANDREI A. BUCKAREFF
Marist College

An account of what sort of causal integration is necessary for an agent to exercise agency is offered in support of a soft-line response to Derk Pereboom's four-case argument against source-compatibilism. I argue that, in cases of manipulation, the manipulative activity affects the identity of the causal process of which it is a part. Specifically, I argue that causal processes involving direct manipulation fail to count as exercises of intentional agency because they involve heteromesial causal deviance. In contrast, standard deterministic causal processes do not involve heteromesial causal deviance and are agency-preserving. The upshot is that there is a relevant difference between a causal process involving direct manipulation by another agent and a deterministic causal process that involves no such intervention. If this is right, then Pereboom's four-case argument does not pose a threat to source-compatibilist theories of free will and moral responsibility.

Key Terms: Agency, Intention, Causal Deviance, Manipulation, Free Will, Powers, Causation, Processes

> "It seems to me that issues of freedom and control should not drive one's ontological commitments in action theory. The correct strategy should be to begin with the most plausible account of action and then see how one can resolve questions about … what constitutes the freedom of the agent." –Berent Enç, (2003: 38, n. 44)

1. Introduction

In contrast to leeway-compatibilists, who emphasize possessing the ability to do otherwise in decision-making as a necessary condition for free will, source-compatibilists reject a leeway condition. Instead, source-compatibilists emphasize that what matters for free agency is what occurs in the *actual* process that gives rise to and includes the agent's decision-making. While accounts vary, what is agreed upon is that free will is exercised when an agent satisfies some conditions sufficient to make it true that they are the source of their decision. Following Michael McKenna, I will refer to any such conditions as "the Compatibilist-friendly Agential Structure (CAS)" (2008: 142). Assuming that an agent's being the source of their actions is both a necessary and sufficient condition for exercising free agency, source-compatibilists endorse the claim that causal determinism does not threaten sourcehood.

Contact: Andrei A. Buckareff <andrei.buckareff@marist.edu>

Not surprisingly, source-*incompatibilists* deny that sourcehood is compatible with causal determinism. The most compelling cases for source-incompatibilism have involved the use of manipulation arguments against source-compatibilism. Manipulation arguments have been developed by source-incompatibilists to show that, even if an agent satisfies some CAS conditions for free will, satisfying those conditions is not sufficient for an exercise of agency to be free.

Michael McKenna has helpfully represented the general structure of manipulation arguments as follows (McKenna 2008: 143):

1. If S is manipulated in manner X to A, then S does not A of her own free will and is therefore not morally responsible for A'ing.
2. An agent manipulated in manner X to A is no different in any relevant respect from any normally functioning agent determined to do A from CAS.
3. Therefore, if S is a normally functioning agent determined to A from CAS, she does not A of her own free will and therefore is not morally responsible for A'ing.

Regarding (2), source-incompatibilists maintain that given that both manipulation cases and cases involving apersonal

causal determinism involve deterministic causal processes, they are relevantly similar (Pereboom 2001: 116).

Two types of responses to manipulation arguments have emerged in the past twenty-five years: soft-line replies and hard-line replies. Soft-line responses take aim at premise (2) of the general manipulation argument. Importantly, they underscore the difference between cases where an agent's actions involve apersonal causal determinism and argue that such cases are quite different from those involving manipulation by another agent. Hard-line replies target premise (1) of the foregoing argument. Such responses highlight the similarities between cases of apersonal causal determinism and manipulation cases and argue that manipulated agents satisfy everything required by CAS and, hence, are free and morally responsible (McKenna 2008: 144).

In this paper, I develop a soft-line response to one manipulation argument, namely, Derk Pereboom's four-case argument (2001: 110-117; 2014: 76-80; 2022: 33-37).[1] Moreover, I only focus on one of the four cases he presents, namely, the first case, which involves *direct* manipulation. I argue that the differences between cases of direct manipulation and apersonal causal determinism are quite stark. In fact, I argue that they are so radically different that in the case of the former, not only is *free agency* undermined, but *intentional agency*, more generally. This is because cases of direct manipulation of an agent involve agency-undermining (not just freedom undermining) features that are structurally identical to what we find in some cases of basic causal deviance. If causal

[1] The other most-often discussed manipulation argument is the zygote argument against compatibilism, developed by Alfred Mele (2006). The zygote argument is most like the second of Pereboom's four cases, involving a form of indirect or distal manipulation. I am only interested in proximal or direct manipulation in this paper. That said, strategies like the one taken up here have been offered against the zygote argument. Broadly, in brief, it has been argued that owing to the effective intentions of the manipulator playing an executive role in cases of indirect manipulation (with the manipulated agent executing the plan of the manipulator), there is a significant difference between a case involving indirect manipulation versus natural causal determinism. Examples of such strategies can be found in Barnes (2015), Herdova (2021), Schlosser (2015), and Waller (2014).

processes involving manipulation are isomorphic to certain types of cases of basic causal deviance, then they describe scenarios where an agent's intentional agency is undermined.

I will proceed as follows in this essay. First, I will introduce Pereboom's four-case argument. I will then discuss the similarities between cases involving what John Bishop (1989) christened "heteromesial basic causal deviance" and direct manipulation. Finally, I will argue that, given that heteromesial basic causal deviance undermines intentional agency, *mutatis mutandis,* direct manipulation also undermines intentional agency. Importantly, the features that undermine intentional agency in cases of direct manipulation are missing in cases involving causal determinism by apersonal natural causal processes.

2. Pereboom's Four-Case Argument

Pereboom takes the locus of free will to be in the practical decisions made by agents (2001: xxi; 2007: 86; 2022: 1-2). (Henceforth, I should be understood to be referring to practical decisions when I use the locution "decision" and its cognates.) Pereboom never specifies the nature of decision-making beyond identifying decisions as mental actions (2022, 1). For my purposes here, I will assume that decision-making is a type of intentional mental action whereby an agent effortfully resolves some practical uncertainty and forms an intention as an outcome. Given that Pereboom focuses on decisions made in a context involving "deliberative mechanisms" constitutive of practical reasoning (cf. 2022: 10, 19, 34), I will focus on decision-making that is part of a larger causal process that is the agent's composite intentional mental action of making up their mind about what to do for reasons. I assume that the process in question commences with the agent's acquiring an intention to resolve some practical uncertainty. Forming an intention in such a case is the terminus of the process.[2] I take it that in such a deliberative context a decision is made *by* deliberating. Deciding in this case involves an additional effort on top of the effort exerted in deliberation to settle the question about what to do (just as kicking involves a distinctive type of effort that is in addition to the effort involved in simply moving one's leg). A decision is made in response to pro- and con-reasons for the options that have been considered in deliberation. The reasons in question may be normative reasons that favor one option over another as well as the agent's own motivational states. This has all been very quick, but it should suffice for my purposes.

In the four-case argument, we have an agent who satisfies four compatibilist "causal integrationist conditions" for free and morally responsible agency found in the literature. The cases are supposed to both preserve the intuition that (i) the manipulated agent is exercising intentional agency while (ii) failing to be free and morally responsible (2014: 76).

Pereboom presents the case of Professor Plum, who decides to murder Ms. White and does so. Pereboom assumes that Plum satisfies the following compatibilist criteria that have been separately offered in the literature as individually sufficient for free agency (2001: 100-110):

[2] The account of deciding I assume here borrows from (and departs in various ways) from the positions articulated in Clarke (2010); McCall (1987); Coffman (2022); Mele (2003: chapter 9; 2005b; 2017: chapter 2; and 2022); Miller and Schwarz (2014), and Shepherd (2015). Of these, I have found the account developed in Shepherd (2015) to be the most informative theory that most closely approximates my own views.

A. Ayer/Hume: The action is the unconstrained outcome of motivational states that flow from the character of the agent at the time.
B. Frankfurt: The action issues from first-order motivational states that mesh with an agent's higher-order motivational states directed at the first-order states being effective in action.
C. Fischer and Ravizza: The action is the product of the activity of a moderately reasons-responsive mechanism (where reasons include a range of practical reasons, including moral and egoistic considerations).
D. Wallace: Plum possesses the powers of reflective self-control ("the power to grasp and apply moral reasons" and "the power to control or regulate his behavior by the light of such reasons") required for an agent to be held morally accountable (Pereboom 2001: 110).

Pereboom presents four cases involving the decision to murder White by Plum. Each of the four cases described by Pereboom "exhibit varying ways in which Plum's decision to kill White might be *causally determined* by factors beyond his control" while satisfying the four compatibilist conditions Pereboom outlined (2014: 76, emphasis added; cf. 2022: 35).

Case 1: A team of neuroscientists can remotely manipulate Plum's neural states. Via such neural interventions "the manipulators enhance Plum's disposition to reason self-interestedly at the requisite time, so that they know that as a result it is causally ensured that he will decide to murder White and he will want to so decide" (Pereboom 2014: 76). They do this by simply pressing a button immediately before he commences deliberating about his situation. "Plum would not have killed White had the neuroscientists not intervened, since his reasoning would not have been sufficiently egoistic to produce this decision" (Pereboom 2022: 34-35).

Case 2: Plum was created by neuroscientists, only they do not directly manipulate him, but he was programmed to reason in ways that are often (but not exclusively) rationally egoistic. Egoistic reasons causally determine his action.

Case 3: Plum was not created by nefarious neuroscientists, but was "determined by the rigorous training practices of home and community so that he is often but not exclusively rationally egoistic" (Pereboom 2001: 114). Egoistic reasons deterministically cause his murdering White.

Case 4: Plum is an ordinary human living in a world where causal determinism is true and he murders White for egoistic reasons.

Again, in each case, what is assumed is that Plum satisfies all four of the compatibilist conditions that are assumed to be individually sufficient for him to exercise free agency and be morally responsible for his decision to kill White.

Pereboom claims that if we are compelled to deny the freedom and moral responsibility of Plum in Cases 1-3, then we should deny that Plum is responsible in Case 4. Pereboom denies that the causal determination of Plum's decision *by other agents* is a relevant difference. For instance, Pereboom argues that if things were exactly the same in Case 1 and Case 2, and if the agent had been manipulated by a spontaneously generated machine with no intelligent designer, then Plum would lack moral responsibility (2001: 115; 2014: 79).

Pereboom's argument can be rendered as follows (see 2014: 79):

P1. Plum is not free and morally responsible in Case 1.

P2. There are no differences between Cases 1 and 2, 2 and 3, and 3 and 4 that can explain in a principled way why Plum would not be free and morally responsible in the former of each pair but would be in the latter.
C1. So (P1) and (P2).
P3. If (C1), then Plum is not free and morally responsible in Case 4.
C2. So Plum is not free and morally responsible in Case 4.

Pereboom claims, in defense of what I have presented as premise (P2) that "[t]he salient factor that can plausibly explain why Plum is not responsible in all of these cases is that in each he is causally determined by factors beyond his control to decide as he does" (2014: 79).

In the remainder of this paper, I will aim at showing that premise (P2) of Pereboom's argument is false, and, hence, premise (2) of the general manipulation argument constructed by McKenna is false.[3]

3. Basic causal deviance and manipulation: Working out the similarities

In this section, I will argue that in Case 1 of Pereboom's four-case argument, we have a situation identical to what we find in certain types of cases of causal deviance in the literature on the causal theory of action (CTA).

The CTA is often described as the standard story of intentional action in the philosophy of action (see Velleman 1992 and Aguilar and Buckareff 2010). In broad outline, the general commitments of the CTA can be schematized as follows:

(CTA) Some activity *A* (whether overt or mental) of an agent *S* is an intentional action if and only if it is caused in the right way and causally explained by some appropriate rationalizing mental item(s).

That some activity must be "caused in the right way" is a necessary condition for it to count as an intentional action owing to worries about causal deviance. *Basic* causal deviance is of particular concern.

In cases of basic causal deviance, an agent's activity is caused by some appropriate mental cause (e.g., a belief-desire pair, an intention, a volition) that represents that activity as either its goal or a means to an outcome. The difference between a case of basic causal deviance and a case when a mental cause produces some activity in the right way lies in the

[3] The soft-line solution I am offering here reflects a strategy similar to the soft-line replies offered by Mele (2005a), Mickelson (2010), and Deery and Nahmias (2017). I depart from them by focusing on how cases of direct manipulation involve a form of basic causal deviance (none of them raise this problem). Also, unlike these other authors, I *explicitly* rely on a causal realist metaphysics of causation (on which causal processes involve constellations of manifesting reciprocal causal powers together *producing* outcomes) and a causal theory of intentional agency that builds upon the assumed theory of causation. Mele and Mickelson do not explicitly endorse any particular metaphysics of causation. Deery and Nahmias endorse James Woodward's (2003) interventionist theory of causal *explanation.* Woodward (2015) has explicitly denied that his concerns are ontological/metaphysical but are, rather, methodological. *Qua* methodological proposal, his theory of causal explanation is concerned with "how we should reason about various important concepts in the scientific enterprise (such as 'cause')" (ibid.: 3578). The account is ultimately a pragmatic one where a particular model of causation is given as a means to achieve the goals of scientific investigation. Insofar as Deery and Nahmias rely on Woodward's interventionism, they are doing something very different from what I am doing here. My project is explicitly metaphysical and relies on a set of specified metaphysical assumptions. They appear to be engaged in a more methodological/explanatory project. The projects are, arguably, complementary.

causal pathway the mental cause takes to produce the intended activity. Differently stated, in cases of basic causal deviance, "the deviance *affects the causal link between mental states and basic action*" (Bishop 1989, 133). Perhaps the most famous case is Donald Davidson's nervous climber.

> A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and want might so unnerve him as to cause him to loosen his hold. (Davidson 1973/2001: 79)

The pathway in this case is deviant or wayward because the causal work done by the rationalizing mental causes is mediated by the climber's being unnerved.

The challenge posed by basic causal deviance has led proponents of the CTA to modify the theory in various ways. The problem seems to stem, at least in part, from the mental causes of some activity playing a merely ballistic role. In the wake of the challenge, CTA proponents have offered different accounts of what is required for some activity to be caused "in the right way" by the relevant mental causes.[4] Most CTA defenders assign a guiding and sustaining causal role to irreducible intentions whose contents represent plans that match the unfolding activity. The details of these various accounts are not important for my purposes here.

What I hope to show is that just as causal processes that include some activity involving basic causal deviance fail to count as exercises of intentional agency, so also activity that involves manipulation of the sort that we have in Case 1 fails to involve any intentional agency by Plum1. This is because cases of direct manipulation such as we find in Case 1 involve basic causal deviance. Specifically, the failure to exercise intentional agency is owing to the causal process involving *heteromesial causal deviance*.

John Bishop coined the term *heteromesial causal deviance* to pick out causal processes that run "from intention to matching behavior through the intentional actions of a *second* agent" (1989, 125). There is some controversy over whether *all* heteromesial causal processes undermine the intentional agency of an agent.[5] That said, there is widespread agreement that causal processes that involve what John Bishop identifies as *preemptively* heteromesial causal deviance fail to count as exercises of intentional agency (see, e.g., Bishop 1989; Brand 1984; Enç 2003; Peacocke 1979). (Henceforth, for ease, in what follows, I will simply use "heteromesial causal deviance" to pick out cases of preemptively heteromesial causal deviance.)

Perhaps the best-known example of heteromesial causal deviance in the literature was offered by Christopher Peacocke. Peacocke asks his readers to imagine "a knowledgeable neurophysiologist who decides on a particular occasion to produce in me exactly the motor impulses needed to realize what he knows, from my neurophysiological states, to be my intentions" (1979: 87). In Peacocke's case, "the chain from intention to bodily movement [passes] through the intentions of a second person" (ibid.). Such cases are taken to block the agent's executive control and undermine their agency and moral

---

[4] For some representative responses, see Aguilar (2010); Bishop (1989); Brand (1984(; Enç (2003); Mele (1992 and 2003); Stout (1996; 2002; 2005; 2007; 2012); and Thalberg (1984).

[5] For instance, Brand (1984) takes all such cases to be agency undermining, while Bishop (1989) allows that not all such cases are a threat.

responsibility. This is so because, "it is wholly up to the neurophysiologist to decide how my body is to move, and [. . .] this fundamental feature of the situation remains even when he adopts a general policy of matching my own intentions" (Bishop 1989: 161). All cases of heteromesial causal deviance involve a wayward causal pathway that goes from an agent's intention and passes through the intention of another agent. Importantly, the second agent has the power to ensure that what they intend to happen will, in fact, be realized. As Bishop notes, "the second agent is part of a system that provides the first agent with, at best, only *indirect* control over the movements of [their] body" (Ibid.: 159). It is the pathway through the intention of another agent with the power in question that is taken to make such causal processes agency-undermining.

While cases of heteromesial causal deviance in the literature may typically involve the intentional action of another human agent who intends to achieve a particular outcome as an element in the process, that any such case involves the activity of another agent is not essential. Suppose a system that is not an agent in the sense of "agent" with which we are concerned in the philosophy of agency (e.g., a fail-safe device implanted by a nefarious neurosurgeon) were programed to execute the intentions of the first agent when they align with the intentions of the second agent that designed the system. If such a system were to intervene in the causal process that would otherwise count as an exercise of intentional agency, then the intervention of the non-human system would still suffice for heteromesial causal deviance and undermine the status of the causal process as counting as an exercise of intentional agency. The reason, simply, is that the system would be implementing the intentions of its designer by proxy.

How are cases of heteromesial causal deviance and direct manipulation like we have in Case 1 alike? Both involve the activity of a second agent who intentionally determines an outcome by directly intervening in a neural process to guarantee an outcome. Consider Case 1 again described as a case of heteromesial causal deviance. Suppose that Plum1 strongly desires to kill White and, hence, is disposed to kill White. But he has some doubts about whether his total reasons favor killing White and he wants to be confident that it would be in his best interest to identify with said desire and intend to kill White. Plum1 desires to make up his mind about whether to kill White and believes he must do so immediately. Forthwith, Plum1 acquires an intention to make up his mind whether to kill White. The team of nefarious neuroscientists know about the neural states of Plum1, including those that constitute his strong desire to kill White and his intention to make up his mind about what to do. When Plum1 commences deciding what to do, they deterministically manipulate the constitutive neural states of Plum1's egoistic desires to ensure that he will acquire the intention to kill White (Pereboom 2014, 76). The process concludes with him identifying with his desire to kill White, acquiring the intention to kill White.

Notice that the foregoing is structurally identical to Pereboom's Case 1. Plum1 wants to make up his mind about what to do. Plum1 intends to make up his mind and commences the process of deliberating and deciding what to do. The nefarious neuroscientists are aware of the neural realizers of Plum1's practical uncertainty and his intention. They intervene in the process of decision-making by directly manipulating the constitutive neural activity of Plum1's making up his mind. They do this to ensure that Plum1 will execute his intention to make up his mind, and they deterministically direct the process so that it will conclude with Plum1's acquiring the intention to kill White.

Case 1 is structurally isomorphic to a standard case of heteromesial causal deviance. Thus, it is not a stretch to conclude that any case of manipulation that is like Case 1 *involves* heteromesial causal deviance. In both this case and Peacocke's classic case of heteromesial deviance involving overt activity we have a scenario with the following features. In both cases we have the first agent intending to do something. The second agent detects the neural activity constitutive of the first agent's intention and the second agent intends that the first agent will execute their intention in a specific way. The second agent knows that by directly manipulating the neural mechanisms that realize the process, the outcome the second agent intends will in fact be realized (see table 1). Given the isomorphism between the two cases, if heteromesial causal deviance undermines something that happens from counting as an intentional action and, hence, as an exercise of intentional agency, then cases of direct manipulation, such as Case 1, involve something occurring that fails to count as an exercise of intentional agency.

Table 1: Comparison of Peacocke's heteromesial deviance and Pereboom's direct manipulation

Agent *S* intends to perform bodily movement *A*	Plum1 intends to decide whether or not to kill White
Neuroscientist intends that *S* will *A*	Neuroscientists intend that Plum1 will intend to kill White
Neuroscientist causes neural activity in S's motor cortex knowing that this will result in *S*'s *A*-ing	Neuroscientists cause neural activity in Plum1 that renders him strongly disposed to reason egoistically, knowing that doing so will result in Plum1 acquiring the intention to kill White
S performs *A*	Plum1 acquires the intention to kill White

Before explaining what it is about these sorts of cases in virtue of which Plum1's intentional agency, and not just his free agency, is undermined, it is worth considering a few objections to my reasoning thus far.

3.1. Objection 1: Matching Intentions

I can imagine a defender of the four-case argument responding at this point that Plum1 does not intend to kill White at the beginning of the process (he intends to decide whether or not to kill White). They may point out that, on the other hand, in Peacocke's scenario, the neuroscientist's intention matches the intention of the agent. But the neuroscientists in Case 1 do not have the same intention as Plum1 (they intend that he will intend to kill White). Therefore, Case 1 cannot be a case of heteromesial causal deviance.

In response, nothing about a case of heteromesial causal deviance requires that the external agent's intention exactly match the intention of the agent who initiates the causal process. Consider the following case. Suppose that Maria intends to donate to Oxfam. Mary is an intervening neuroscientist who intends that Maria donate to a famine relief organization (it could be any organization) and acts to ensures this outcome. The contents of their intentions are not the same. Maria's intention represents a plan that involves her donating to Oxfam. Mary's intention represents a plan that involves her ensuring that Maria will donate to a famine relief organization. While their intentions are not the same, it is still a case of heteromesial causal deviance.

My interlocutor may still not be convinced. They may argue that, in the case of Plum1 and the neuroscientists, Plum1 intends to make up his mind and the neuroscientists intend that he will intend to kill White. Thus, their intentions are quite different. Plum1 is practically uncertain and intends to settle on whether to kill White. The neuroscientists intend that Plum1 acquire the intention to kill White.

The difference in the intentions of Plum1 and the neuroscientists is irrelevant. The causal pathway from Plum1's intention to make up his mind to his acquiring the intention to kill White still goes through the intentions of the neuroscientists who intend that Plum1's mind will be made up in a particular way and intervene in the neural processing that realizes his decision-making to guarantee their desired outcome. This is no less a case of heteromesial causal deviance.

3.2. Objection 2: Who's in Control?

My interlocutor may point to a difference between the sort of case Peacocke mentions and the case we are considering involving Plum1. Bishop describes what occurs in Peacocke's scenario as one where,

> once the first agent's intention to do *a* is formed, [a] it is entirely up to the second agent to initiate efferent neural stimulation in the first agent's brain in order to produce bodily movements intrinsic to *a*-ing. [b] The first agent exercises no further control once the intention has been formed, and so the *a*-ing behavior cannot count as the first agent's own genuinely *basic* action. (1989: 159)

It may be argued that the neuroscientists in Case 1 are not controlling the process of Plum1's acquiring the intention to kill White. Rather, they are taking measures to ensure that Plum1 decides to kill White. Their causal contribution is part of the total cause of Plum1's decision. While in the case Bishop describes, the neuroscientist is controlling the behavior.

In response, we have a scenario with both of the features Bishop identifies as agency-undermining. I will take them up in reverse order. Regarding (b), in Case 1, the nefarious neuroscientists are causally preempting Plum1's coming to intend differently by directly manipulating specific neural states of Plum1's once he intends to make up his mind, guaranteeing the outcome of Plum1 acquiring the intention they intend, thereby vitiating his control. As for (a), that Plum1 even acquires an intention is entirely up to the neuroscientists given their power to manipulate his neural states. And that Plum1 finally intends as he does depends upon their stimulating his neural states in order to produce their intended outcome. Thus, we have a causal process that is under the control of the neuroscientists. Hence, Plum1's agency is undermined.

3.3. Objection 3: Heteromesial Pathways are Not Necessarily Deviant

My imaginary interlocutor may note at this point that not all causal processes involving heteromesial causal pathways involve basic causal deviance. They would be correct. Suppose that, owing to neural damage in his dorsolateral prefrontal cortex, a reliable prosthetic device has been implanted in the brain of Plum1 that runs proxy for the damaged neural pathways in his brain. The device enables him to make decisions. The device is activated whenever Plum1 is in a state of practical uncertainty that he intends to resolve and enables Plum1 to make up his mind about what to do. While causal processes involving such a device would take a heteromesial causal pathway, the activation of the prosthetic does not undermine Plum1's agency in this case. Rather, the device *enables* Plum

to make decisions. The role of the neuroscientists could be like the prosthetic, ensuring an outcome without there being any causal deviance.

There is an important difference between such a prosthetic and the neuroscientists in Case 1 (or, for that matter, a computer that is exercising control that is similar to the control exercised by the neuroscientists). In the case involving the neuroscientists, the process of acquiring an intention is ultimately one that is guided (either in whole or in part) by the intentions of the neuroscientists. Plum is not aware of the neuroscientists, and any potential responsiveness to feedback would not be a cooperative venture. The neuroscientists are not simply aiding Plum 1, faithfully helping him to execute his intention to make up his mind. The neuroscientists (or any functionally equivalent computer) are guaranteeing an outcome that they want and intend to bring about. The prosthetic has no intentions. It is not an agent. It is an artifact that was designed to occupy a functional role within the system that is the agent. As such, it is functionally integrated into that system. It has been assigned a causal role in the agent that would otherwise be played by an organic component of the agent.

If we suppose that the neuroscientists are simply aiding Plum1 in realizing what he wants or intends to do, having pledged themselves to always do that which will enable Plum1 to achieve *his* goals (whatever their own actual desires and goals might be), then the neuroscientists are like the reliable prosthetic. In such a case, Plum1 and the neuroscientists would share responsibility for the decision to kill White (see Bishop 1989: 161-162). But this is *not* the sort of case Pereboom envisages.

3.4. Coda

If I am right, then Case 1 presents a scenario involving heteromesial causal deviance. As such, we have a scenario where Plum1 is not only failing to exercise *free* agency, but he is not exercising *intentional* agency. Pereboom explicitly states that Case 1 involves the neuroscientists manipulating Plum1 "at the neural level, but with the result that his mental states and actions feature the psychological regularities and counterfactual dependencies characteristic of genuine agency" (2014, 76). If I am right, Case 1 does *not* involve genuine agency. Owing to the intervention of the neuroscientists, we have a heteromesial causal process that undermines the agency of Plum1. *It is a mistake to describe Plum1 as having performed the intentional mental action of deciding to kill White.*

Suppose that I am right about Case 1 involving heteromesial causal deviance that undermines intentional agency. The incompatibilist can just dig in their heels, concede that, in Case 1, Plum1 does not exercise intentional agency and, therefore, in Case 4, Plum4 does not exercise intentional agency. Such incompatibilists may be inspired by the work of Helen Steward (2012) on agency incompatibilism. Steward argues that exercising agency should be understood in terms of settling whether it will be the case that *p.* She argues that the truth of causal determinism implies that it is already settled that *p.* The incompatibilist who follows Steward may then argue that Case 1 and Case 4 are relevantly similar insofar as they involve scenarios where it is already settled that *p* prior to the agent making up their mind. Thus, they may argue that Pereboom has shown (however inadvertently) that causal determinism undermines intentional agency.

I will not take up a discussion of the arguments offered by Helen Steward for agency incompatibilism here.[6] Rather, I will assume (with Pereboom) that causal determinism is not a threat to *intentional* agency and show that Case 1 and Case 4 are relevantly different. While Case 1 involves a form of basic causal deviance that is widely regarded as undermining intentional agency, Case 4 does not obviously present us with a scenario involving Plum4's intentional agency being undermined.

4. Why We Have a Breakdown of Intentional Agency in Case 1 But Not in Case 4

The success of Pereboom's argument hinges on the claim that there is no relevant difference between Case 1 and Case 4. I will show that Case 1 and Case 4 are quite different and the intuition that Plum in Case 1 is not free and morally responsible rests on Plum's failure to exercise intentional agency owing to the manipulation by the neuroscientists resulting in a different kind of causal process than what we have in cases of actual intentional agency. That the causal process is causally deterministic is irrelevant.

An assumption that will guide how I proceed in what follows is that an adequate response to Pereboom cannot be neutral on questions about the metaphysics of intentional agency and the metaphysics of causation. While the schema for CTA provides conditions for intentional *action* it fails to provide us with a general account of intentional *agency*. A theory of intentional agency should allow us to account not only for agency in acting, but also exercising agency in intentionally omitting to act (where the omission is a basic omission) and in bringing about outcomes over which an agent lacks direct control. I will assume that it is best that we understand intentional agency as a causal process that includes, in the case of action, both the acquisition of the proximal mental cause(s) of an action and the outcome of the action as components. How we formulate a more general causal theory of intentional agency will depend upon how we think about causal processes. A few words about my assumed framework are in order.[7]

I assume a causal realist account of causal processes on which at least some of the properties possessed by objects (including those of complex objects, such as human agents) are irreducible causal powers/dispositional properties. On this view, a causal process exhibits a unity and directedness inherited from the constellation of manifesting reciprocal causal powers of its interacting constituent components.[8] Each causal power active in a causal process is directed at manifestations activated by interactions with other reciprocal causal powers. Outcomes of causal processes are the polygenic products of constellations of reciprocal manifesting causal powers constitutive of the process. What *type* of causal process

[6] For responses to Steward's agency incompatibilism, according to which intentional agency, generally, is incompatible with causal determinism, see Beebee (2014); Bishop (2014); Boxer (2013); Clancy (2013); Clarke (2014); Garnett (2013); and Levy (2013).

[7] In many respects, my own work is motivated by worries raised in Frankfurt (1978) for the causal theory of action. For discussion of the importance of Frankfurt's work, see Aguilar (2020). Accounts that bear a family resemblance to the theory I am offering are proposed and defended in Hyman (2015); Mantel (2018); Setiya (2007); and Stout (2002; 2006; 2007; 2010; 2012).

[8] For more on this sort of picture of causation, see Buckareff (2017), Chakravartty (2005 and 2007), Heil (2012), Marmodoro (2017), Molnar (2003), Mumford and Anjum (2011), and Williams (2019).

any *token* causal process will count as is fixed by what objects and their powers are active in the process.

A causal process inherits a *telos* from its constitutive manifesting powers and terminates once the outcome at which it is directed is achieved. Importantly, if additional causal powers are introduced into a causal process, then the type of causal process of which it is a token will have changed. Just consider what happens when someone is being poisoned and an antidote that masks the lethal causal powers of the poison is given to the person to whom it was given. With the introduction of the antidote, the identity of the causal process has changed from a poisoning to the neutralizing of a poison — more colloquially, the process becomes a failed attempt at a poisoning. The causal process is different from one where a poison is given to someone without any antidote being administered. But differences in causal processes need not involve a difference in the goal of the processes. Consider the neutralization of hydrochloric acid. You can achieve the same goal by using sodium bicarbonate or sodium hydroxide. The causal process that is neutralizing hydrochloric acid that involves using sodium bicarbonate is of a different type from the one involving sodium hydroxide, but both processes share the same *telos*.

The general causal theory of intentional agency (CTAg) I assume builds on the foregoing powers-in-process metaphysic of causation.[9] I assume that agents are best understood as functionally integrated systems of fundamental objects and their causal powers. Capacity-ascriptions that we make to complex objects (including human agents) are made true by constellations of causal powers of objects. Their causal powers that make them distinctively agents are together directed at being manifested in response to reasons and producing purposeful behavior. I assume that an exercise of agency is best understood as a causal process characterized by the manifestations of the agent's reciprocal causal powers constitutive of their motivational states, proximal intention, and executive capacity for responding to practical reasons that are collectively directed at achieving the agent's goal in that circumstance as represented in the content of their proximal intention. An exercise of intentional agency commences with an agent's acquiring a proximal intention to act or a proximal intention to achieve some goal that will result from acting or omitting to act in a certain way. The acquisition of an intention is in response to some of the agent's motivational states and their consideration of normative reasons. An exercise of intentional agency concludes either once the agent's intended goal is achieved or the requisite means that the agent believes will render the agent's goal have been realized.

The following represents the core commitments of CTAg schematically.

(CTAg) For any causal process *A*, *A* is an exercise of intentional agency by an agent *S* if and only if (i) *A* commences with the non-accidental and endogenous acquisition and manifestation in response to reasons by *S* of an intention *I* (and the constitutive causal powers thereof) to *O* (where *O* may be a tokening of an action-type, an omission, or the consequences of an action or omission); and (ii) *S* guides the process through to the final execution of *I* by the ongoing manifestation of *S*'s executive capacity to sensitively respond to the constellation of inputs from various manifesting reciprocal causal powers of

[9] The general framework I present here is presented with varying degrees of detail in Aguilar and Buckareff (2015 and 2022); Buckareff (2011; 2018; and 2022); and Buckareff and Kasper-Buckareff (2014).

both *S* and objects in *S*'s environment to which *S* endogenously responds until *S* successfully executes *I*.

I will not defend CTAg here. For my purposes in what follows, I will refer to the first condition, (i), as *the endogenous initiation condition* and the second condition, (ii), as *the endogenous guidance condition*.

CTAg presents us with a causalist theory of intentional agency that (a) can be applied to account for all the ways we manifest intentional agency; (b) is consistent with the CTA; and (c) is more informative than the generic statements of CTA often found in the literature. For those who do not wish to buy into the metaphysical framework in which CTAg is embedded, I would invite them to fill in the background with whatever other alternative they like. But the account will need to be informative enough to be able to offer a principled way to individuate process types—in particular, causal processes that count as exercises of intentional agency and those that do not. The metaphysical framework I assume and how it is applied in CTAg provides a principled way of distinguishing token causal processes that are exercises of agency and those that are not (tokenings of process types are identified and individuated by the types of objects and powers active in a causal process). To see why this is the case, some words are in order about each of the conditions specified in CTAg.

First, regarding (i), *the endogenous initiation condition*, the causal process that is an exercise of intentional agency is initiated by the endogenous acquisition of an intention (and manifestation of the constitutive powers of the intention) in response to the reciprocal powers constitutive of considered reasons and motivational states that favor an outcome. With respect to (ii), *the endogenous guidance condition*, notice that the causal process is guided by the agent through to the execution of the intention and the agent does so via the manifestation of an executive capacity of theirs being manifested in response to interacting with various causal powers of the agent, the constitutive powers of practical reasons to which they respond, and objects in the agent's environment to which the agent is endogenously responding. A causal process that satisfies (i) and (ii) will involve a constellation of reciprocal powers of the agent and objects in the agent's environment mutually manifesting in response to their interacting with one another producing an intended polygenic outcome.

Importantly, the process-type of which an agent's exercise of agency is a tokening will be determined by the causal powers that are activated as part of the process. Any addition or subtraction of relevant causal powers will result in a different process. In particular, the activation of causal powers of an agential system that are alien to the system that is the agent under consideration will render the process type being tokened one that does not count as an exercise of intentional agency. This is particularly so when the agent is not knowingly cooperating with another agent. Therefore, cases of heteromesial causal deviance fail to count as exercises of intentional agency and this is also why a manipulation case like Case 1 counts as a case of heteromesial causal deviance. Let me explain.

Consider Case 1 again. The *endogenous initiation condition* is satisfied. We have the endogenous acquisition of an intention by Plum to decide whether to kill White and the process does not involve any alien causal powers. But *the endogenous guidance condition* is not satisfied. The process is not endogenously guided by Plum1. It is exogenously guided by the nefarious neuroscientists. The process is ultimately guided by the neuroscientists who manipulate Plum1's neural states so that he will be disposed to assign greater weight to

egoistic reasons knowing that by doing this they will ensure that Plum1 will acquire the intention to kill White and subsequently execute his intention. The type of process that unfolds is not one that can be identified as an exercise of intentional agency by Plum1. Plum1 is not *the* originator or source (in a compatibilist sense of "source" or "origin") of the intention to kill White that is acquired. And the activity of the neuroscientists is not to aid Plum1 in exercising his agency (to use David Pears' language (1975, 67), they are not the functional equivalent of an "action-aid"). Rather, they intend to manipulate Plum1's decision-making process to guarantee their intended outcome. Thus, following Peacocke (1979, 88), given the intention and power of the neuroscientists to guarantee that Plum will form and execute the intention to kill White, if anyone is *controlling* the formation of Plum1's intention to kill White, it is the neuroscientists and not Plum1.

But what about the deterministic nature of the causal processes in Case 1 and Case 4? Isn't that the most relevant similarity? Whether or not the causal process is deterministic is completely irrelevant to whether it counts as an exercise of intentional agency.[10] Suppose that the causal process in Case 1 is one in which the outcome is not necessitated but merely rendered more or less likely given the constellation of powers active in the process.[11] That the process does not involve the necessitation of the final outcome is irrelevant. The only thing that undermines the agential status of the causal process that is Plum1's manipulated (ersatz) decision-making is the manifestation of the executive causal powers of the neuroscientists. It is the manifestation of these alien causal powers in the causal process that make it false that the process of Plum1's coming to acquire an intention to kill White counts as the tokening of any kind of intentional agency by Plum1.

If I am right, then, *contra* Pereboom, there is a relevant difference between Case 1 and Case 4. Case 1 does not involve an exercise of intentional agency because it involves a tokening of the wrong type of causal process. But in Case 4, Plum4 does exercise intentional agency. That Plum4's activity in Case 4 was a deterministic causal process is irrelevant for its status as an exercise of intentional agency. For those unconvinced, some more explanation of the relevant difference between Case 1 and Case 4 is in order.

In Case 1, while the system that is Plum1 is integral to the causal process that occurs, the causal powers of Plum1 *qua* agent are only part of the total constellation of causal powers of *agents* that are manifested in the process. More importantly, the causal powers constitutive of the executive capacities of Plum1 are not manifesting in cooperation with the powers of some other agents directed at a common goal. Rather, the manifestation of the executive capacities of Plum1 is in the service of the goals of other agents. We have a causal process of Plum1 coming to acquire an intention to kill White that cannot be truthfully described as Plum1 endogenously manifesting his executive capacity to directly guide the process of making up his mind. Rather, it is one that is better described as the neuroscientists manifesting their collective capacity to control Plum1. So, if I am right, then (P2) in the argument I offered to represent Pereboom's reasoning is false. Regarding (P2),

[10] See Mele (2005a), where the same general point is made.

[11] Mumford and Anjum (2011 and 2015) develop a powers-in-process account of free agency on which, given the additive and subtractive effect of the constellation of powers, the process of decision-making is not deterministic and the outcome is a disposition to act in a certain way that does not necessitate what an agent finally does.

there is a relevant difference between Case 1 and Cases 2-4. Case 1 involves heteromesial causal deviance and, hence, is a causal process that does not count as an exercise of intentional agency. Given this, since Plum1 does not exercise intentional agency in Case 1, he is not free and morally responsible in Case 1. But Case 4 does not present the same difficulties. Cases 1 and 4 are anisomorphic. The agency undermining features of Case 1 are not present in Case 4. Hence, given the relevant difference between the cases, we have not been given a reason to take Case 4 to involve any elements that would undermine the free agency and moral responsibility of Plum4.

I can imagine my interlocutor demanding more reasons for agreeing with my conclusion. To those I now turn.

First, as mentioned above, Case1, like other cases of heteromesial causal deviance, involves a failure to satisfy *the endogenous guidance condition.* Following Robert Audi, I take it that an exercise of intentional agency under the guidance of an agent is "a response to, not a mere effect of, a reason" (1993: 177). A "guiding idea here is that [an exercise of agency] one performs for a reason [is] *one's own*" (Ibid.: 164). Of course, Plum1 wanted to resolve his practical uncertainty and intended to make up his mind about what to do with respect to White for reasons; but when Plum1 commences making up his mind, his mind is made up *through* the action of the neuroscientists who ensure that he weighs his reasons in a particular way. As mentioned above, the process involves exogenous determining influence by another system that is guiding the process. Therefore, the process is not one involving Plum1 guiding his own deliberation and deciding in response to his own reasons. The process is guided by the neuroscientists who manipulate him to weigh reasons in a particular way that will necessitate his acquiring the intention to kill White.

Second, the failure to satisfy the *endogenous guidance condition* is, in part, owing to a breakdown in the system that is Plum1. The mereological sum of Plum1 and the neuroscientists is not a causally and functionally integrated system that we can identify as a collective agent with Plum1 and the neuroscientists as parts. While they are loosely identifiable as one system, owing to the control the neuroscientists can and do exert over Plum1, we have a system whose function is not to pursue the goals of Plum1 or to pursue the jointly agreed upon goals of Plum1 and the neuroscientists. By coupling Plum1 to another deliberative system in this way, as Berent Enç notes about cases of heteromesial causal deviance, the functional integrity of Plum1 is no longer retained (2003: 130). This is owing, in part, to the fact that the unfolding causal process is not sensitive to the content of Plum1's intention but rather to the shared intention of the neuroscientists that takes causal precedence over Plum1's (Stout 2005: 96-97; 2007: 150-152). Any feedback in the causal process ultimately loops back to the neuroscientists. The buck stops with them, not Plum1. The process that commences is no longer Plum1's own given that he is in a subservient role to the neuroscientists whose manipulation of Plum1 is exerted as a means to achieving *their* intended goal. To use the language preferred by Rowland Stout, what Plum1 does belongs to "the operation of a teleological mechanism" that belongs to the neuroscientists, not Plum1. There is agency here. But it is the neuroscientists who are exercising agency, not Plum1, since "the agency belongs to [whomever] embodies the teleological potentiality the realisation of which results in the behaviour" (Stout 2007: 151).

It helps to contrast the foregoing with what happens in Case 4. Causal determinism does not result in any change in the identity of the causal process that is identical with Plum4's exercise of agency. In considering Case 4, we can hold everything fixed from Case 1

with the exception of the intervention of the neuroscientists in the causal process. All that we add in Case 4 is that the relevant causal processes involve causally necessary connections between the causing and outcomes. In Case 4, Plum4 would satisfy both *the endogenous initiation condition* and *the endogenous guidance condition.* But, again, with Case 1, the constellation of reciprocal manifesting causal powers active in the causal process is anisomorphic to what we find in Case 4, resulting in a failure to satisfy *the endogenous guidance condition.* This is so because, in Case 1, the causal process involves heteromesial causal deviance.

If I am right, then Case 1 and Case 4 are quite different. It is not the determinism of the causal process that matters. Case 1 fails to involve a causal process that counts as exercising endogenous guidance owing to its involving heteromesial causal deviance. Case 4 does not involve heteromesial causal deviance and involves a causal process that counts as an exercise of intentional agency. Therefore, in Case 1, Plum1's putative decision fails to count as an exercise of intentional agency while Plum4's decision-making in Case 4 *is* an exercise of intentional agency. If this is right, then a manipulated agent like Plum1 is relevantly different from Plum4, and the difference is of significance for their statuses as free and morally responsible agents (since you cannot have an exercise of *free* agency if you fail to have an exercise of *intentional* agency). If they are relevantly different, then premise (P2) of the argument that represents Pereboom's reasoning is false.

It may be argued that the anisomorphism between the causal processes in Case 1 and Case 4 is only apparent or at least that the activity of the neuroscientists is not as relevant for the purposes of determining whether we have a genuine exercise of intentional agency. Pereboom asserts that "[a]gency is regularly preserved in the face of certain involuntary momentary external influences" (2014: 76). He writes that:

> Finding out that the home team lost can cause one to reason and behave more egoistically and less charitably, and news of winning a prize stands to make one reason and act more generously, but the conditions of agency remain intact. We commonly suppose that acting on such influences is compatible with moral responsibility, but we can imagine an egoism-enhancing momentary influence that preserves agency but does preclude responsibility. (Pereboom 2014: 76)

Pereboom's description of the cases belies the differences. Importantly, the case he describes is perfectly consistent with exercising endogenous control in both initiating and guiding the causal process that is one's exercise of intentional agency. The agent is responding to features of their environment. In discovering the home team lost, one's deliberative process that follows is neither the result of nor guided by the intentional agency of an intervening party that intends that a specific outcome is generated and knows that their intervention will result in that outcome. Moreover, *the endogenous initiation condition* and *the endogenous guidance condition* can be satisfied when one is causally influenced in the way Pereboom describes. But such a case is quite different from what we have in Case 1. Most significantly, the case of learning that one's home team lost does not involve heteromesial causal deviance (or any other species of basic causal deviance).

Conclusion

I have aimed to show that cases of *direct* manipulation are relevantly different from cases involving apersonal causal determinism. Manipulation cases like we have in Case 1 involve agency-undermining heteromesial causal deviance. Given this feature of such cases, they

are quite different from cases involving causal determinism that do not involve any such deviance in the causal process.

I have intentionally remained mostly quiet about cases of *indirect* manipulation such as we find in Case 2 and Case 3 of the four-case argument and Alfred Mele's (2006) zygote argument. I have not taken up such cases given that they are quite different from Case 1 in the four-case argument. I am not sure about whether a soft-line strategy is best for responding to such cases. I am quite sympathetic to the line of reasoning taken up by some proponents of soft-line strategies.[12] That said, I also find the hard-line approach defended by Michael McKenna (2008) to be quite attractive. But, as I see it, what makes Case 1 of the four-case argument importantly different is that we are presented with a case where the manipulation undermines the *intentional* agency of the manipulated agent. If cases involving indirect manipulation involve the agency of the manipulated agent being compromised, they are scenarios that appear to involve the undermining of the agent's *free* agency. Can Pereboom restructure the argument starting with Case 2? I suppose. But I think the intuitive force of his argument rests on his presentation of Case 1. I would hypothesize that the intuition that the agent is not free in Case 1 is owing to the scenario being one where there is no intentional agency being exercised.

Regardless of what a future line of reasoning from manipulation cases like we find in the four-case argument might look like, I believe I have shown that Pereboom is mistaken in asserting that there is no relevant difference between Cases 1 and 4. There is a relevant difference, and it is a significant difference. The threat posed to source-compatibilism by the four-case argument is, therefore, a chimera.

Acknowledgements: Earlier versions of this paper were presented to audiences at the 10th European Congress of Analytic Philosophy, the Fifth Workshop on Agency in the Mountains, the 2023 Pacific Division Meeting of the American Philosophical Association, the 2023 Canadian Philosophical Association Congress, and to the philosophy departments at the University of Manitoba and Bilkent University. I benefitted immensely from feedback from commentators and members of the audience on each occasion. I am especially grateful to Jesús H. Aguilar, Mark Balaguer, Gerry Beaulieu, John Bishop, Taylor W. Cyr, Matthew Flumer, Alex Grzankowski, Rory Harder, Michael McKenna, Adam Murray, Alyssa Ney, Jonathan D. Payton, Rebekah Rice, T. Raja Rosenhagen, Joshua Shepherd, Chris Tillman, Bill Wringe, and two anonymous referees for this journal for helpful feedback on earlier versions of this essay.

References:

Aguilar, Jesús H. (2010). Agential Systems, Causal Deviance, and Reliability. In Jesús H. Aguilar and Andrei A. Buckareff (Eds.), *Causing Human Actions: New Perspectives on the Causal Theory of Action* (85-100). The MIT Press.

Aguilar, Jesús H. (2020). The standard story of action and the problem of agential guidance. *Crítica, 52*(155), 3-25.

Aguilar, Jesús H. and Andrei A. Buckareff (2010). The Causal Theory of Action: Origins and Issues. In Jesús H. Aguilar and Andrei A. Buckareff (Eds.), *Causing Human Actions: New Perspectives on the Causal Theory of Action* (1-26). The MIT Press.

[12] I find the strategies adopted by the authors mentioned in footnote 1 to be quite promising.

Aguilar, Jesús H. and Andrei A. Buckareff (2015) A Gradualist Metaphysics of Agency. In Andrei A. Buckareff, Carlos Moya, and Sergi Rosell (Eds.), *Agency, Freedom, and Moral Responsibility* (30-43). Palgrave-Macmillan.
Aguilar, Jesús H. and Andrei A. Buckareff (2022). Agency and causation. In Luca Ferrero (Ed.), *Routledge Handbook of Philosophy of Agency* (27-36). Routledge.

Audi, Robert (1993). Acting for reasons. In Robert Audi, *Intention, Action, and Reason* (145-178). Cornell University Press.
Barnes, Eric Christian (2015). Freedom, Creativity, and Manipulation. *Noûs, 49(3),* 560-588.
Beebee, Helen (2014). Radical Indeterminism and Top-Down Causation. *Res Philosophica, 91*(3), 537-545.
Bishop, John (1989). *Natural Agency: An Essay on the Causal Theory of Action.* Cambridge University Press.
Bishop, John (2014). Causal Pluralism and the Problem of Natural Agency. *Res Philosophica, 91*(3), 527-536.
Boxer, Karen (2013). Assessing the argument for agency incompatibilism. *Inquiry, 56*(6), 583-596.
Brand, Myles (1984). *Intending and Acting: Toward a Naturalized Action Theory*. The MIT Press.
Buckareff, Andrei A. (2011). How Does Agent-Causal Power Work? *The Modern Schoolman: Special Issue on Free Will and Moral Responsibility, 88*(1-2), 105-21.

Buckareff, Andrei A. (2017). A Critique of Substance Causation. *Philosophia, 45*(3), 1019-1026.
Buckareff, Andrei A. (2018). I'm Just Sitting Around Doing Nothing: On Exercising Intentional Agency in Omitting to Act. *Synthese, 195*(10), 4617-4635.
Buckareff, Andrei A. (2022). Two-Way Powers as Derivative Powers. In Michael Brent and Lisa Maracchi Titus (Eds.), *Mental Action and the Conscious Mind* (228-254). Routledge.
Buckareff, Andrei A. and Lara Kasper-Buckareff (2014). Legal Agreements and the Capacities of Agents. In Enrique Villanueva (Ed.), *Law and the Philosophy of Action* (195-220). Rodopi.
Chakravartty, Anjan (2005). Causal Realism: Events and Processes. *Erkenntnis, 63*(1), 7-31.
Chakravartty, Anjan (2007). *A Metaphysics for Scientific Realism: Knowing the Unobservable.* Cambridge University Press.
Clancy, Sean (2013). A Strong Compatibilist Account of Settling. *Inquiry, 56*(6), 653-665.
Clarke, Randolph (2010). Making up One's Mind. In Joseph K. Campbell, Michael O'Rourke, and Harry S. Silverstein (Eds.), *Action, Ethics, and Responsibility* (67-84). The MIT Press.
Clarke, Randolph (2014). Agency and Incompatibilism. *Res Philosophica, 91*(3), 519-525.
Coffman, E. J. (2022). Unsettledness and the Intentionality of Practical Decisions. *Philosophical Explorations, 25*(2), 220-231.
Davidson, Donald (1973/2001). Freedom to Act. In Donald Davidson, *Essays on Actions and Events,* 2nd ed. (63-81). Oxford University Press.

Deery, Oisín. and Eddy Nahmias (2017). Defeating Manipulation Arguments: Interventionist Causation and Compatibilist Sourcehood. *Philosophical Studies, 174*(5), 1255-1276.
Enç, Berent (2003). *How We Act: Causes, Reasons, and Intentions*. Oxford University Press.
Frankfurt, Harry (1978). The Problem of Action. *American Philosophical Quarterly, 15*(2), 157-162.
Garnett, Michael (2013). Freedom and Unpredictability. *Inquiry, 56*(6), 666-680.
Herdova, Marcela (2021). The Importance of Being Ernie. *Thought: A Journal of Philosophy, 10*(4), 257-263.
Heil, John (2012). *The Universe as We Find It.* Oxford University Press.
Hyman, John (2015). *Action, Knowledge, and Will.* Oxford University Press.
Levy, Neil (2013). Are We Agents at All? Helen Steward's Agency Incompatibilism. *Inquiry, 56*(4), 386-399.
Mantel, Susanne (2018). *Determined by Reasons: A Competence Account of Acting for a Normative Reason.* Routledge.
Marmodoro, Anna (2017). Aristotelian Powers at Work: Reciprocity Without Symmetry in Causation. In Jonathan Jacobs (Ed.), *Causal Powers* (57-76). Oxford University Press.
McCall, Storrs (1987). Decision. *Canadian Journal of Philosophy, 17*(2), 261-288.
McKenna, Michael (2008). A Hard-Line Reply to Pereboom's Four-Case Manipulation Argument. *Philosophy and Phenomenological Research, 77*(1), 142-159.
Mele, Alfred R. (1992). *Springs of Action: Understanding Intentional behavior*. Oxford University Press.
Mele, Alfred R. (2003). *Motivation and Agency.* Oxford University Press.
Mele, Alfred R. (2005a). A Critique of Pereboom's "Four-Case Argument" for Incompatibilism. *Analysis, 65*(1), 75-80
Mele, Alfred R. (2005b). Decisions, Intentions, and Free Will. *Midwest Studies in Philosophy, 29*(1), 146-162.
Mele, Alfred R. (2006). *Free Will and Luck.* Oxford University Press.
Mele, Alfred R. (2017). *Aspects of Agency: Decisions, Abilities, Explanations, and Free Will.* Oxford University Press.
Mele, Alfred R. (2022). Are Practical Decisions Mental Actions? In Michael Brent and Lisa Maracchi Titus (Eds.), *Mental action and the conscious mind* (255-271). Routledge.
Mickelson, Kristin (2010). The Soft-Line Solution to Pereboom's Four-Case Argument. *Australasian Journal of Philosophy, 88*(4), 595-617.
Miller, Jeff and Wolf Schwarz (2014). Brain Signals Do Not Demonstrate Unconscious Decision-Making: An Interpretation Based on Graded Conscious Awareness. *Consciousness and Cognition, 24*, 12-21.
Molnar, G. (2003) *Powers: a study in metaphysics.* Oxford University Press.
Mumford, Stephen and Rani Lill Anjum (2011). *Getting Causes from Powers.* Oxford University Press.
Mumford, Stephen and Rani Lill Anjum (2015). Freedom and Control: On the Modality of Free Will. *American Philosophical Quarterly, 52*(1), 1-11.
Peacocke, Christopher (1979.) *Holistic Explanation.* Oxford University Press.
Pears, David (1975). The Appropriate Causation of Intentional Basic Action. *Crítica, 7*(20), 39-72.
Pereboom, Derk (2001). *Living Without Free Will.* Cambridge University Press.

Pereboom, Derk (2007). Hard Incompatibilism. In John M. Fischer, Rorbert Kane, Derk Pereboom, and Manual Vargas, *Four Views on Free Will* (85-125). Blackwell.
Pereboom, Derk (2014). *Free Will, Agency, and Meaning in Life.* Oxford University Press.
Pereboom, Derk (2022). *Free Will.* Cambridge University Press.
Schlosser, Markus. (2015) Manipulation and the Zygote Argument: Another Reply. *Journal of Ethics, 19*(1), 73-84.
Setiya, Kieran (2007). *Reasons without rationalism.* Princeton University Press.
Shepherd, Joshua (2015). Deciding as intentional action: control over decisions. *Australasian Journal of Philosophy, 93*(2), 335-352.
Steward, Helen. (2012) *A Metaphysics for Freedom.* Oxford University Press.
Stout, Rowland (1996). *Things that Happen Because They Should: A Teleological Approach to Action.* Oxford University Press.
Stout, Rowland (2002). The Right Structure for a Causal Theory of Action. *Facta Philosophica, 4*(1), 11-24.
Stout, Rowland (2006). *Action.* Acumen.
Stout, Rowland (2007). Two ways to understand causality in agency. In Anton Leist (Ed.), *Action in Context* (137-153). de Gruyter.
Stout, Rowland (2010). What are You Causing in Acting? In Jesús H. Aguilar and Andrei A. Buckareff (Eds.), *Causing Human Actions: New Perspectives on the Causal Theory of Action* (101-113). The MIT Press.
Stout, Rowland (2012). Mechanisms that Respond to Reasons: An Aristotelian Approach to Agency. In Fran O'Rourke (Ed.) *Human Destinies: Philosophical Essays in Memory of Gerald Hanratty* (81-97). University of Notre Dame Press.
Thalberg, Irving (1984). Do Our Intentions Cause Our Intentional Actions? *American Philosophical Quarterly, 21*(3), 249-260.
Velleman, J. David (1992). What Happens When Someone Acts? *Mind, 101*(403), 461-481.
Waller, Robyn Repko (2014). The Threat of Effective Intentions to Moral Responsibility in the Zygote Argument. *Philosophia, 42*(1), 209-222.
Woodward, James (2003). *Making Things Happen: A Theory of Causal Explanation.* Oxford University Press.
Woodward, James (2015). Methodology, Ontology, and Interventionism. *Synthese, 192*(11), 3577-3599.
Williams, Neil (2019). *The Powers Metaphysic.* Oxford University Press.