



Causalism Without Causation

Carolina Sartorio¹

Accepted: 8 May 2023 / Published online: 15 June 2023
© The Author(s), under exclusive licence to Springer Nature B.V. 2023

Abstract

Moore's *Mechanical Choices* is ripe with interesting ideas. Here I'll focus on a particularly intriguing one that intersects with some aspects of my own work. It's the suggestion that causalism should be amended in a way that doesn't require causation. At first, this suggestion may sound absurd: How can causalism survive without causation, of all things? But I think that Moore is actually right about the main suggestion. I don't think he's right for the right reasons, but he's still right about the main idea. So, the aim of this paper is to explain how causalism can survive without causation, and how it may not.

Keywords Agency · Causalism · Causation · Michael Moore · Moral responsibility

1 Introduction

Michael Moore's *Mechanical Choices* (Moore 2020) is ripe with interesting ideas. Here I'll focus on one that I found particularly intriguing and that intersects with some aspects of my own work. It's the suggestion that causalism (the standard view of agency, and of the kind of control that underlies attributions of agency and moral responsibility) should be amended in a way that doesn't require causation (this is mostly covered in chapter 11 of Moore's book, but also in parts of chapters 3 and 7).

At first, this suggestion may sound absurd: How can a view like causalism survive without *causation*, of all things? But I think that Moore is actually right about the main suggestion. I don't think he's right *for the right reasons*, but he's still right

✉ Carolina Sartorio
carolina.sartorio@rutgers.edu

¹ Rutgers University, New Brunswick, USA

about the main idea. So, the aim of this paper is to explain how causalism can survive without causation, and how it may not.

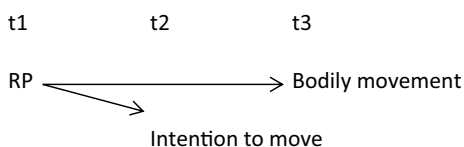
2 Moore's Deflated Causalism

Moore's main motivation for proposing a "deflated" version of causalism—a causalism that doesn't require causation—is the hope that this would let us reconcile our agency and moral responsibility with some recent findings in neuroscience by Libet and others (Libet et al. 1983). I'll focus on a particular challenge to which these findings are said to give rise: *the epiphenomenal challenge*, as Moore calls it, which seems to raise the most trouble for causalism as a theory of agency, as well as for our intuitive sense of agents as beings who can be in control and can be morally responsible for our behaviors.

On its most extreme version, the epiphenomenal challenge is the challenge that our intentions¹ are epiphenomenal, or that they don't have the causal powers that we attribute to them in thinking that they cause our bodily movements when we act intentionally. The claim that our intentions have such causal powers is intimately tied to causalism, the standard and most widely accepted view of agency, as that view is typically understood in terms of the thesis that intentional actions are behaviors that are caused (in the right way) by mental states or events such as intentions, beliefs and desires, etc.

The epiphenomenal challenge is motivated as follows. The neuroscientific findings in question (those presented in, e.g., Libet et al. 1983) allegedly show that there is some unconscious brain activity that inevitably precedes the formation of an intention: the "Readiness Potential" or RP. The RP is then said to start a causal process that results in the bodily movement *without going through the agent's intention*, as represented by the following diagram:

The Epiphenomenal Challenge



This type of causal structure is called an "epiphenomenal fork." The intention is an epiphenomenon in the sense that it's itself causally inert (as indicated by the absence of an arrow linking the intention to the bodily movement); however, it's caused by something that causes the bodily movement: the RP, which acts as a common cause.

As Moore notes, a common reply to this challenge has been to reject the claim that the neuroscientific results show that intentions are in fact epiphenomenal. For

¹ Moore uses "willings" but I'll stick with "intentions," which I think is a more neutral and less ambiguous term. (What are willings supposed to be: are they tryings, decisions, or something else altogether? It's not fully clear. Mele raises these questions in his 2017: 72–3.).

example, some have argued that this interpretation of the findings relies on a picture of agency that is way too crude—in particular, one that is blind to the fact that the formation of an intention can be a more complex process that begins some time before the agent’s conscious awareness of that process (see, e.g., Mele 2009). Thus, the fact that the agent isn’t yet aware of the presence of the intention doesn’t entail that the process of forming an intention hasn’t already started.

However, Moore wants to offer a more ambitious reply. Moore’s reply accepts, for the sake of the argument, the proposed interpretation of the results, and thus, that our intentions are epiphenomenal. But Moore argues that this is compatible with our retaining the capacity to exercise our agency and moral responsibility, at least in some cases. Moore also argues that this type of reply doesn’t contradict causalism—or what he sees as the “essence” of causalism. For, he argues, causalism doesn’t actually require that our intentions be causally efficacious. More provocatively put: Moore argues that *causalism doesn’t require causation* (by the relevant mental events or states).²

What could the essence of causalism be, if not one that requires causation? Moore seems to think that it’s a form of *means-end control*. Arguably, the following statement captures what he has in mind:

Moore’s deflated causalism: We perform actions to the extent that we behave in ways that are controlled by our intentions, in the general way means can be used to achieve certain ends. Although this kind of control is typically causal, it doesn’t *have* to be, for some non-causal forms of control can do the required work.³

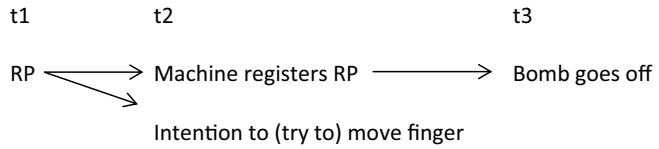
For Moore, we know that the type of control that we exercise is non-causal if the means that we use temporally follow the ends. This isn’t because Moore is assuming that backwards causation is metaphysically impossible (in chapter 11 he says that this is an issue he wants to sidestep; see p. 420). Rather, it’s because he thinks that we (ordinary human beings who lack the capacity to time-travel) are not able to engage in backwards causation. As a result, if we’re able to exercise control over some ends by using means that temporally succeed the ends, it must be because that control isn’t causal.

Moore illustrates his view with an example of the following kind (I’m simplifying the example a bit while remaining true to the basic structure; see his discussion of the “Paralyzed Patriot” case in chapters 3 and 11):

Paralysis: An agent, A, cannot move his index finger due to its being paralyzed, but he can intend to move it (or, at least, to try to move it). A has been hooked up to an interface machine that exploits the recent neuroscientific findings in the following way: the machine can read the brain activity that inev-

² Moreover, this isn’t just because of what Moore thinks about omissions (Moore thinks that, whereas causalism should be able to capture the type of agency involved in omitting to act, omissions are not the kinds of things that can be involved in causal relations). But it’s also for this reason, as I explain in Sect. 5 below.

³ Here I am paraphrasing Moore (see his p. 74 and 423).

Paralysis

itably precedes A's intention (the RP), and when the machine registers that activity, it directly starts a process that culminates in a bomb going off. (Say, a button is pushed—one that an unparalyzed being would have been able to push by using his index finger—by the machine itself, and the pushing of the button sets off the bomb.) A knows all of this, so, wanting the bomb to go off, he forms the relevant intention. The bomb goes off as planned.

Moore would argue that in circumstances like this A is responsible for the bomb going off, as the bomb's going off was within A's control. But A's intention to move his finger (or to try to move it) didn't cause the bomb to go off.

The way Moore sees it, the causal structure of this case is another epiphenomenal fork, of the following kind:

A's intention is an epiphenomenon. For the common cause, the RP, causes both the bomb's going off and the intention; the intention itself doesn't cause the bomb to go off.

Moore argues that cases like this show that being involved in an epiphenomenal fork is compatible with acting, with being in control, and with causalism (or, at least, with its essence). For, he would argue, the agent in this case retains all that matters even if his intentions are not causally efficacious. In particular, the relevant means-end connection that his deflated causalism requires exists: A has control over whether the bomb goes off by controlling whether he intends his finger to move. But that control isn't causal, because A only has that kind of control thanks to the *non-causal* control he has over whether the RP occurs, in the first place. This is a *backtracking* form of control: A controls whether the RP (a past event) occurs by controlling what he intends to do. In other words, A has control over what's, by then, in the past!

To his credit, Moore seems to recognize how surprising and counterintuitive his suggestion that we can have backtracking control sounds (see, e.g., p. 420). However, he believes that it's worse to let agents in cases of this type off the hook (see, e.g., p. 74 and pp. 428–9). Also, Moore's view is that backtracking control is only possible in this case due to the assumption that the relevant past event (the RP) is, as he puts it, *strongly necessitated* by the willing. (As Moore notes, without this restriction the view has incredibly implausible consequences involving past events that we most definitely don't control; see pp. 426–8.)

Moore doesn't have a fully worked out account to offer about the relation of "strong necessitation," and leaves it at an intuitive level. But he seems to have in mind a kind of physical or biological necessity (certainly not logical or metaphysical necessity). He seems to be thinking: if we accept the scientific findings, we believe

that, as a matter of biology or neurophysiology (or, more generally, natural law), the formation of an intention is inevitably preceded by the relevant unconscious brain activity (the RP). So, Moore's idea is that the paralyzed agent can rely on that strong necessitation relation to exercise backtracking control over the occurrence of the RP, and thus, ultimately, over the final outcome. But, in other cases where that strong necessitation relation doesn't exist, agents lack that kind of control.

3 Against Backtracking Control

I disagree with Moore's assessment about **Paralysis**, and in this section I'll explain where I think the reasoning goes wrong.

First, though, let me note that it's unclear that, even if Moore's deflated causalism could be made to work, it would give us a satisfying form of compatibilism. In particular, I worry that a possible implication of Moore's proposal is that we couldn't be morally responsible *unless* we exploited the neuroscientific findings in the way exemplified by **Paralysis**. For, if we didn't exploit those findings, there is the threat that the chains of events would then be too "deviant" for us to be morally responsible for any outcomes in those chains.

Let me explain. Philosophers of action commonly note that, when a causal chain is deviant (in the sense that it's abnormal or unexpected in some significant way), this is enough to undermine the agent's control (see, e.g., Mele 2017: Sect. 3.4.1). For example, imagine that Martians are secretly monitoring your brain processes. When you form the intention to move your arm, they intervene by forcing your arm to move. In that case, your intention to move causes your body to move (by causing the Martians to intervene, which causes the bodily movement) but it does so in an unexpected or abnormal way. As a result, you're not in control of your bodily movement and you're not responsible for it.

Similarly, then, imagine that the picture Moore is imagining is right and the way in which our intentions are linked to outcomes is by means of an unexpected backtracking connection—one that goes through the RP that occurs prior to each intention. Then the worry is that the chain of events would end up being too deviant. For, while we normally assume that our intentions result in outcomes in a relatively straightforward way, on this proposal they would in fact only do so in an indirect, not purely causal, and quite unexpected kind of way. Could we then be responsible for any outcomes in such ordinary situations? It's unclear that we could. And if we could not, then it seems that we haven't made much progress in answering the heart of the epiphenomenal threat.

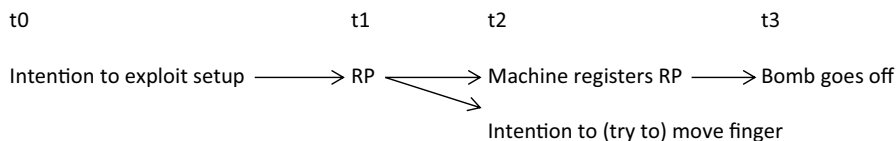
Fortunately, it seems to me that there is an easy way out of Moore's puzzle, and one that doesn't commit us to any form of backtracking control. The basic idea is this: the only reason we're tempted to blame A (the paralyzed agent) for the explosion in **Paralysis**, even though we're assuming (for the sake of the argument) that A's intention to move his finger is causally inert, is that, by the time A forms that intention, he already has a plan that involves *an earlier intention*. The plan is to

exploit the relevant neuroscientific findings as well as what he knows about the machine in order to make the bomb go off. And this plan crucially involves the earlier intention to later form the intention to move his finger so as to set off the causal process that will culminate in the bomb going off.⁴ But note that, in that case, what makes A responsible for the explosion is the earlier intention, not the intention to move his finger. That earlier intention causes the RP to appear, and the relevant process to get started, in accordance with the plan. A's responsibility for the bomb's going off can then be traced back to his responsibility for having formed that earlier intention, which causally resulted in that outcome in the expected way.

But couldn't Moore argue that this reasoning is flawed because Libet's results also show that the earlier intention *itself* is causally inert (since there is another RP that precedes *it*, which does the causal work)? However, this response would fail. Note that I'm only going along with the assumption that the later intention is causally inert for the sake of the argument. And what I'm suggesting is that the only reason we still think A is responsible in that case is that we're also implicitly assuming that some prior intention of A *is* causally efficacious. If it turned out that none of A's intentions were causally efficacious, then I think it would be clear that A is not responsible for the outcome at all.

In other words, what I'm suggesting is that the reason we think the agent in **Paralysis** is responsible for the outcome (even though the intention to move his finger is causally inert) is that we're assuming that the case has a more complex causal structure than the one Moore suggests. That more complex structure can be represented by the following diagram: When the agent forms the intention to exploit the setup, he triggers a causal process that is, by assumption, a bit unusual (as it goes through the RP and not through the intention to move the finger itself). However, the agent can still be responsible for that process, and for the outcome of that process, as he specifically intended for things to happen in precisely that way, given the special circumstances he knew he was placed in.

Paralysis*



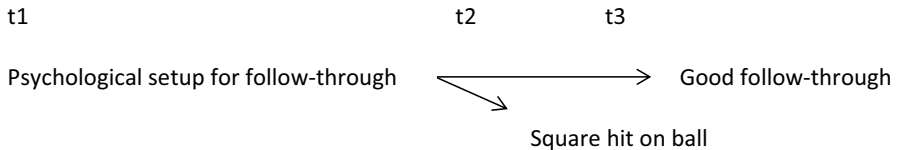
In the relevant respects, this new causal structure resembles another one discussed by Moore in chapter 11 (pp. 420–22): a case of a golfer who achieves a square hit on a golf ball in an indirect kind of way, by aiming for a good follow-through instead of directly aiming for a square hit on the ball (the example is originally from Hornsby 1980). Moore's interpretation of the case, which I'll accept here, is that

⁴ Interestingly, Moore seems to implicitly recognize the existence of the earlier intention when he writes, about his Paralyzed Patriot case: "*Intending to convey the information that he has* (that the British are coming by boat), PP wills the movement of his paralyzed finger..." (pp. 422–3, my emphasis).

the follow-through is epiphenomenal. That is, the follow-through doesn't cause the square hit; instead, a common cause (an earlier act or intention by the golfer, which is the setting herself up for the follow-through), plays the relevant causal role in this case.

The following diagram represents the situation, as Moore is imagining it:

Golfer



Note that here the control is causal, since it traces back to a causally efficacious intention. This is so even if the intention is not the intention to hit the ball in a specific way, but the intention to achieve a good follow-through. For the (unusual) content of the intention involves the best, and rather indirect, way the agent knows of bringing about the desired outcome.

As we have seen, **Paralysis** seems to have a similar structure: an intention by the agent with a special content causes the bomb to go off, and in the intended way. Here too, the (unusual) content of the intention involves the best, and rather indirect, way the agent knows to bring about the desired outcome, given the special circumstances he's in. Thus, the control in **Paralysis** is causal too.

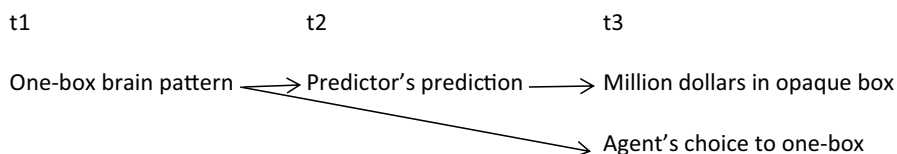
Let me mention one more case discussed by Moore in chapter 11: Newcomb's case. In Newcomb's case, a player is presented with two boxes and is given the choice to take one box or both boxes. One box is opaque, and it contains either \$1,000,000 or \$0; the other box is transparent, and it visibly contains \$1000. The player is told that the content of the opaque box was determined yesterday by a reliable predictor ("the Predictor"), who put \$1,000,000 in it if he could predict that the player would pick only the opaque box, or \$0 if he could predict that the player would pick both boxes. By the time the player gets to make the choice, the Predictor has already made his prediction and the content of the opaque box has already been set in stone (either the money is there or it isn't). In the literature on this problem, the choice to pick both boxes is called the "two-boxing" strategy, and the choice to pick only the opaque box is called the "one-boxing" strategy.

In his book, Moore argues that Newcomb's case has a similar structure to **Paralysis**, and argues on that basis that we should one-box (pp. 423–5). This is Moore's reasoning: when the agent in Newcomb's case gets to choose between one-boxing and two-boxing, his choice gives him non-causal backtracking control over the earlier act by the Predictor. This is because, by assumption, there is a relation of strong necessitation between the agent's choice and some events that occurred prior to that: the prediction by the Predictor and the antecedent facts on which that prediction was based. In particular, Moore is assuming that the agent's choice to one-box is inevitably preceded by some antecedent brain activity, such as a pattern of blood flow in a

certain region of the agent’s brain (one that indicates the intention to one-box), and that the Predictor uses that fact to make his prediction. Given this relation of strong necessitation, Moore thinks that the agent has non-causal control over the Predictor’s prediction because he has non-causal control over the antecedent activity in his own brain. (This is in the same way that the agent in **Paralysis** has non-causal control over the explosion because he has non-causal control over the occurrence of the antecedent brain activity, the RP.) And, again, Moore’s suggestion is that the relevant notion of control is non-causal and backtracking: there is no time-travel involved, and the agent doesn’t causally influence the Predictor’s earlier choice. The Predictor has already made his choice, based simply on the evidence that he had at that past time.

The following diagram represents how Moore seems to be thinking about the case:

Newcomb



Notice that this is another epiphenomenal fork, where the agent’s choice to one-box is an epiphenomenon. The common cause (the agent’s brain activity at t1) both causes the Predictor’s prediction at t2 and the agent’s choice at t3. Moore claims that the agent’s choice at t3 gives him non-causal control over the prior events at t1 and t2, and thus over the outcome of getting the million dollars.

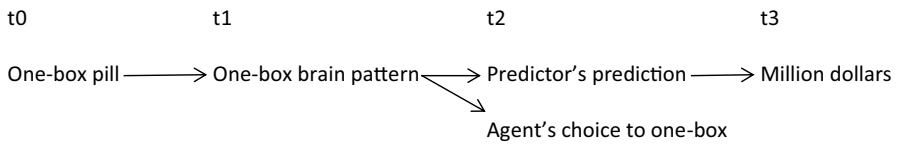
Now, of course, we (two-boxers) know that one-boxing is irrational. It’s irrational because either the money is already in the box or it isn’t, and the agent making the choice at t3 has no control over those past events. There is no backtracking control (for us, ordinary human agents without the capacity to time-travel). Therefore, two-boxing is the only rational option.

As a result, as two-boxers also note, the only way in which we could get rich in a Newcomb case would be if we could somehow turn ourselves into irrational agents (agents who have the disposition to one-box) earlier on (see, e.g., Joyce 1999: 154). That way the Predictor would know that we have such dispositions and would put the million dollars in the opaque box, and we would then proceed to (irrationally) one-box.

Imagine, for example, that we had in our possession a “one-box pill”: a pill that would force us to make the irrational choice to one-box at t3 by generating the relevant activity in our brain at an earlier time. We would then get the million dollars. But, of course, in that case what gives us control over the million dollars is not the act of one-boxing itself, but the earlier act of taking the pill. And the control that we have in that case is, again, causal.

Note that, in imagining this variant of the case, we have effectively turned the case into an analogue of the **Paralysis** and **Golfer** cases, one that has the following structure:

Irrational Newcomb



The only main difference between **Irrational Newcomb** and the other cases that share this same structure is that here the agent must turn himself into an irrational being in order to exploit the setup, whereas he doesn't have to do any such thing in **Paralysis** or **Golfer**.

4 Lewis and Backtrackers

In the previous section I argued against Moore's solution to the epiphenomenal challenge. In this section I'll connect that discussion with Moore's discussion of Lewis and backtrackers.

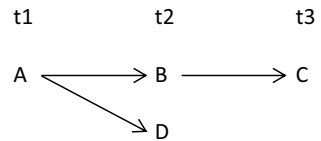
Moore suggests that one main reason the idea of backtracking control has such a "bad rap" is due to Lewis's work on causation (Lewis 1986a). In that now classical paper Lewis argued that the best theory of causation—a counterfactual theory—shouldn't make room for backtrackers (at least not in ordinary situations where agents lack time-travel abilities). But Moore thinks that the same negative attitude against backtrackers isn't warranted if one is interested in a broader notion of control, one that allows for *non-causal* forms of control.

As we will see in the next (and final) section, I think Moore is right that we should be interested in a broader notion of control, one that allows for non-causal forms of control. And I also agree with Moore that this is consistent with the essence of causalism. But what I certainly disagree with Moore about is his suggestion that there are *backtracking* non-causal forms of control (for ordinary human beings like us, who lack the capacity to time-travel). In fact, I believe that Lewis's argument about causation can be expanded into an argument against Moore's views on control.

Let me explain. As noted above, Moore embraces a theory of control that can be supported by strong necessitation relations. And, as it turns out, such a theory of control is analogous to the type of theory of causation that Lewis was trying to distance himself from when he put forth his counterfactual theory of causation: the so-called "regularity theories" of causation. Both theories—regularity theories of causation and Moore's theory of control—are based on robust necessitation relations. And Lewis's argument against regularity theories (and in favor of counterfactual theories) used, precisely, epiphenomenal forks. Lewis used epiphenomenal forks because he thought that those cases illustrate the fact that we need certain *asymmetrical* resources to account for the asymmetries of causation. These resources, he argued, are the asymmetries of counterfactual dependence, which disallow backtracking (on the intended reading of counterfactual dependence which counterfactual theories of causation use).

To illustrate, consider this general causal structure of an epiphenomenal fork:

Epiphenomenal Fork



Lewis noted that structures of this kind make trouble for regularity views of causation. For in this case, by assumption, D doesn't cause C. However, there could still be a regularity (a strong necessitation relation) that allowed us to backtrack from D to A (and then from A to B to C). This is what would happen if, given the laws and some of the actual circumstances, D couldn't have been caused by anything other than A. Under those conditions, a regularity view of causation would wrongly imply that D caused C. In contrast, a counterfactual theory (one that doesn't allow backtracking) avoids this result. For the counterfactual "If D hadn't occurred, C wouldn't have occurred" is false according to a standard, non-backtracking reading. This is because, if we're not allowed to backtrack, we must assume that, if D hadn't occurred, A would still have occurred, which would have then caused C.

Now, Lewis may not be fully right about causation. But it seems to me that what Lewis says about causation applies, even more plausibly, to *control* more generally. We (again: ordinary human beings without the capacity to time-travel) don't have control over the past. Thus, epiphenomenal forks of this kind (where there is a backtracking relation of strong necessitation between D and A) can be used to argue against a view of control, like Moore's, that allows us to backtrack. In other words, I would suggest that the epiphenomenal forks that Moore uses as *illustrations* of his view of control should more plausibly be seen as supporting an *argument against* his views on control (or against any other view of control based on strong necessitation relations that allow for backtracking).

Let me also comment on how I think this is tied to another topic involving Lewis and backtrackers: Lewis's classical response to the consequence argument for incompatibilism. The consequence argument is an argument for the incompatibility of determinism and the ability to do otherwise (and thus free will, if one thinks free will requires the ability to do otherwise; see van Inwagen 1975 and, also, 1983). Roughly, the argument goes as follows: if determinism is true, and thus the present is a necessary consequence of the remote past and the laws, then we are powerless over what we do (we don't have any alternative possibilities of action in the present). For we don't have the ability to render the remote past or the laws false, and from this powerlessness over the past and the laws our powerlessness over the present also follows.

Lewis (1981) replied to this argument by noting that, in a sense that doesn't require backwards causation or the capacity to perform miracles, we may in fact have the capacity to render the conjunction of the laws and the past false. The sense in which we may have this capacity is simply this: (assuming certain compatibilists are right and we are, in the relevant sense, able to act otherwise despite determinism

being true) we are able to do something that is such that, *had we done it, the conjunction of the laws and the past would have been different*. On the assumption of determinism, this counterfactual must be true, on pain of logical contradiction.

Now, is this a backtracking counterfactual that gives us control over the past? Not for Lewis, because Lewis argues that what we have to imagine as being different is mainly the laws (although we have to imagine that the laws would have been different not in a generalized but in a localized way; this is his “local miracle” compatibilism). Now, Lewis also argues that (to avoid more generalized miracles or violations of laws) we have to imagine that the immediate past (not the remote past) is different. However, and this is what’s key for our purposes here, we must not imagine that the immediate past is different in any *specific* way. For doing that would result in a particular counterfactual being true, which would in turn give us control over particular events in the past, which we clearly don’t have. If all we’re imagining, instead, is that something would have been different, but it’s undetermined what, then no particular backtracking counterfactual is true, and thus we avoid the unwanted conclusion that we can control specific events in the past (Lewis 1981: 117–8).

Again, I mention all of this not to agree with Lewis on the consequence argument (I don’t think free will requires the ability to do otherwise, and thus I’m not particularly interested in responding to the consequence argument), but only to explain how his answer fits with the rejection of backtracking control. In a nutshell: Lewis would argue that his response to the consequence argument doesn’t commit him to the kind of backtracking control that Moore thinks we can have, and that Lewis would argue we obviously lack.

Here I side with Lewis, then: we *don’t* have control over the past (or, at least: we don’t have control over any *specific* events in the past). But Moore’s views implausibly commit us to that kind of control. According to Moore, the agent who chooses to one-box has control over the Predictor’s prediction, and over the agent’s own antecedent brain activity on which that prediction is based. Similarly, we can have control over the specific brain activity that takes place in our brain prior to the formation of our intentions (if the neuroscientific findings are taken at face value). But we don’t have any such control—in particular, I have argued that Moore hasn’t given us good reason to think otherwise.

5 Other Forms of Non-causal Control

Still, I believe that Moore is right about one main thing: causalism (or, in any case, the essence of the view) *doesn’t* require causation. It doesn’t require causation because the essence of causalism is in fact compatible with non-causal forms of control.

One possible way to see this is to think about potential *non-causal consequences* of our acts. Suppose that you are not paralyzed, and that when you form the intention to move your index finger, your finger moves as a result, and it pushes the button that detonates the bomb. You then have control over a series of events that you can be held responsible for. These arguably include your moving of your finger/your

finger moving, the button being pushed, the explosion, and a victim of the explosion dying. But it also seems plausible to say that you have control overall several non-causal consequences of your act. For example, you non-causally bring about that *some* finger has been moved, that *some* button has been pushed, that *somebody* has died, that the victim's spouse has become a widow/er, etc., and you may be morally responsible for at least some of these consequences.

On certain views of events (such as, again, Lewis's view—presented in Lewis 1986b), these are not genuine events—or, at least, they're not the kinds of events that can enter in causal relations. This is either because they're too disjunctive (such as the state of affairs that consists in somebody dying) or too extrinsic (such as the state of affairs that consists in a person's becoming a widow/er), or because they're more like logical or analytic consequences of causal effects than causal effects themselves (both examples might help illustrate this). But here, again, we don't need to decide this issue. We don't need to determine whether Lewis is right and consequences like these really are non-causal consequences of acts. For the main point I want to make is that there *could* be, at least potentially, non-causal consequences of our acts, and that nevertheless causalism should be able to accommodate them. Surely, causalism is not refuted by the existence (or the possible existence) of non-causal consequences of what we do. Surely, causalism must be interpreted in a way that allows for that possibility.

Now, perhaps one could easily accommodate potential non-causal consequences like these by arguing that causalism is only concerned with the relation between intentions and their more immediate consequences, which are always causal. For example, one could say that causalism is not supposed to accommodate *making a person a widow/er* as a genuine action that the agent performs. The only real actions in this case would be, perhaps, just the basic action of moving the finger, or perhaps also other actions that only involve causal consequences, such as the actions of bringing about the button-pushing, the explosion, and the victim's death. Everything else could be seen as foreseeable consequences of actions, for which you can be responsible in a more derivative way, in the same way we are responsible for other consequences in general.⁵

However, there is another, more important reason why we need to make room for the possibility of non-causal consequences within the causalist framework. It's the fact that our intentional agency includes our *omissions* in addition to our positive actions, and it's an open metaphysical question whether omissions can be causes and effects (assuming omissions are not events but absences of events). For it's an open metaphysical question whether absences in general can be causes and effects.

Moore makes a related point on p. 428. He actually has a view on the topic of omissions, which he has defended in his previous book (Moore 2009): he thinks it's clear that omissions *cannot* enter in causal relations, because they are absences and absences in general cannot enter in causal relations. As a result, Moore thinks that our concept of responsibility must allow for non-causal relations of consequence, if

⁵ I discuss responsibility for non-causal consequences in Sartorio 2022.

it is to accommodate the responsibility involved in omissions. My argument is different. For what I'm arguing is that the question whether omissions can be causes and effects is an *open metaphysical question*, and that the prospects of causalism should not hinge on an answer to such a question. In fact, causalism as a general view of intentional agency should be officially neutral on this question.⁶

And causalism *can* be neutral on this question. For, surely, even if omissions cannot enter in causal relations, they can explain why other things happen (or be part of explanations of why other things happen), and they can be explained by other things. I suggest that that explanatory connection is all that causalism needs. It's an ordinary explanatory connection that obtains between distinct events or states of affairs, one that is typically causal but one that doesn't necessarily have to be causal. (In particular, it wouldn't be causal in the case of omissions if it turned out that omissions cannot be causes and effects.)⁷

For example, imagine that I fail to scratch an itch upon intending not to scratch it, and *because* I intended not to scratch it. Then such an explanatory relation is what makes my omission intentional. This would still be the case if that explanatory connection weren't, strictly speaking, causal. For it seems that we don't need to answer the more basic metaphysical question concerning absence causation to know that the existence of such an explanatory connection is what makes the omission intentional. Thus, the essence of causalism survives, independently of what the right answer to the metaphysical question is.

Philosophers who have written on the metaphysics of causation seem to all agree about this fundamental issue: omissions have explanatory powers, even if they cannot enter in causal relations. In particular, those who argue that omissions cannot enter in causal relations tend to appeal to surrogate (non-causal) concepts that can easily accommodate the explanatory power of omissions. These surrogate concepts include the concepts of quasi-causation (as in Dowe 2001) and causal explanation (as in Beebe 2004 and Varzi 2007). Any of these surrogate relations or concepts could do the relevant work in a causalist account of agency as applied to omissions.

To conclude, let me tie this point back to the main point of the earlier sections. To say that causalism can subsist without causation isn't to say that causalism is free from the past-future asymmetries of ordinary human control. Since I cannot affect the past, my behavior doesn't have any *backtracking* non-causal consequences. But it can have non-causal consequences that are not backtracking, and those are the kinds of non-causal consequences that a deflated causalism should be ready to accommodate in the case of ordinary human beings.

Compare: a time-traveler, or somebody who had the ability to time-travel, would be different from me in that respect. Imagine a counterpart of me who can time-travel. It seems that, *just* by having that ability, my counterpart's behavior has many consequences in the past, especially once one includes her omissions as part of her intentional behavior. For, if she *could have* traveled into the past and causally

⁶ I argue for this in more detail in Sartorio Forthcoming. See also Sartorio 2021.

⁷ It's also an ordinary explanatory connection in the sense that it's not teleological (or irreducibly teleological) in nature, in opposition to what some non-causalist views of agency say.

influenced the past in certain ways, then it seems that a full explanation of the past will have to include her omissive behavior (her not traveling into the past and her not exerting that causal influence). This is so even if she never really did travel into the past. And this is so even if the influence of her omissions isn't causal but quasi-causal (or non-causal).⁸

Acknowledgements I presented an earlier version of this paper at a symposium on Michael Moore's book at Rutgers University in May 2022. Thanks to the organizers, Doug Husak and Alec Walen, and to Michael and the audience members for discussion. Thanks also to Alec for helpful editorial suggestions.

References

- Beebe, H. 2004. "Causing and Nothingness." In J. Collins, N. Hall, and L. A. Paul (eds.), *Causation and Counterfactuals*, Cambridge, Mass: M.I.T. Press, pp. 291-308.
- Dowe, P. 2001. "A Counterfactual Theory of Prevention and 'Causation' by Omission." *Australasian Journal of Philosophy* 79 (2): 216-26.
- Hornsby, J. 1980. *Actions*. Routledge.
- Joyce, J. 1999. *The Foundations of Causal Decision Theory*. Cambridge University Press.
- Lewis, D. 1976. "The Paradoxes of Time Travel." *American Philosophical Quarterly* 13: 145-52.
- Lewis, D. 1981. "Are We Free to Break the Laws?" *Theoria* 47: 113-21.
- Lewis, D. 1986a. "Causation." In *Philosophical Papers*, Volume II, New York: Oxford University Press, pp. 159-213.
- Lewis, D. 1986b. "Events." In *Philosophical Papers*, Volume II, New York: Oxford University Press, pp. 241-69.
- Libet, B., C. A. Gleason, E. W. Wright, and D. K. Pearl. 1983. "Time of Conscious Intention to Act in Relation to Onset of Cerebral Activities (Readiness Potential); the Unconscious Initiation of a Freely Voluntary Act." *Brain* 106: 623-42.
- Mele, A. 2009. *Effective Intentions: The Power of Conscious Will*. Oxford University Press.
- Mele, A. 2017. *Aspects of Agency: Decisions, Abilities, Explanations, and Free Will*. Oxford University Press.
- Moore, M. 2009. *Causation and Responsibility: An Essay in Law, Morals, and Metaphysics*. Oxford University Press.
- Moore, M. 2020. *Mechanical Choices: The Responsibility of the Human Machine*. Oxford University Press.
- Sartorio, C. 2015. "The Problem of Determinism and Free Will Is Not the Problem of Determinism and Free Will." In Mele, A. (ed.), *Surrounding Free Will: Philosophy, Psychology, Neuroscience*, Oxford University Press, pp. 255-73.
- Sartorio, C. 2021. "Responsibility and the Metaphysics of Omissions." In S. Bernstein and T. Goldschmidt (eds.), *Non-Being: New Essays on Non-Existence*, Oxford University Press, pp. 294-309.
- Sartorio, C. 2022. "Responsibility and Causation." In D. Nelkin and D. Pereboom (eds.), *Oxford Handbook of Moral Responsibility*, Oxford University Press.
- Sartorio, C. Forthcoming. *Causalism*. Oxford University Press.
- Sider, T. 2002. "Time Travel, Coincidences, and Counterfactuals." *Philosophical Studies* 110, 2: 115-38.
- van Inwagen, P. 1975. "The Incompatibility of Free Will and Determinism." *Philosophical Studies*, 27, 3: 185-99.
- van Inwagen, P. 1983. *An Essay on Free Will*. Oxford: Oxford University Press.
- Varzi, A. 2007. "Omissions and Causal Explanations." In F. Castellani and J. Quitterer (eds.), *Agency and Causation in the Human Sciences*, Paderborn: Mentis Verlag, pp. 155-67.

⁸ Several philosophers have argued that those with time-traveling abilities could have altered the past (although of course they didn't). They include, again, Lewis (1976), as well as Sider (2002), among many others. I briefly touched on time-travel and omissions in Sartorio 2015.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.