

## CHAPTER 10

# Mental Causation Reconsidered

In most discussions of the problem of mental causation, mental causation is presented as a cause–effect relation between mental and physical items. Mentality and physicality are presented as two sides of a causal exchange. I called this understanding of mental causation the relational understanding of mental causation.

**Relational understanding of mental causation:** mental causation is mental items (events, processes or states) standing in causal relations to physical items (e.g. movements of a person's body).

Philosophers writing about the problem of mental causation are limited to this way of describing what mental causation is, because they assume that 'cause' is an unequivocal term—all causation everywhere is the same kind of thing, so the only thing that can discriminate between different categories of causation is the nature of the relata involved. What is 'mental' about mental causation is that it involves at least one mental relatum. I argued that this understanding of mental causation is a crucial component of the main argument for adopting a physicalist metaphysics of mind. However, it is my view that this is a flawed approach to understanding mental causation.

One of the aims of this book was to explain why the relational understanding of mental causation is presupposed in many debates in philosophy of mind. In the first three chapters, I showed that the relational understanding of mental causation is entailed by a triad of philosophical theories: physicalism, causal theories of intentional action and a relational approach to causation. I argued that, even though these theories are logically independent, in practice they reinforce each other. I called this triad the physicalist triad because the upshot of endorsing these three theories is that physicalism ends up seeming like the only possible metaphysics of mind that stands a chance of saving the phenomenon of mental causation.

---

### How to cite this book chapter:

White, Andrea. 2024. *Understanding Mental Causation*. Pp. 197–204. York: White Rose University Press. DOI: <https://doi.org/10.22599/White.k>. License: CC BY-NC 4.0

My second aim in this book was to try to describe a way to break out of the physicalist triad. In so doing, I hoped to break physicalism's hegemony over our thinking about the mind. The strategy I followed was to focus on what I take to be the weakest element of the physicalist triad, namely its account of human agency. The physicalist triad entails a physicalist/event-causalist description of human agency, where what it is to act is to do something intentionally, and what it is for an action to be intentional is explained in terms of causation by a mental state of the agent, or a mental event involving the agent. And, according to physicalism, these mental items are realised by physical items—most plausibly neural events, or perhaps physical events that are themselves complex and include neural events as parts. The picture of human agency that emerges is a reductive one. What it is for a person to act is nothing more than the triggering of bodily movements by sub-personal events. This picture of human agency is endorsed, at least partially, by Bishop (1989), Brand (1984), Bratman (1987), Dretske (1988), Enç (2003), Mele (1992; 2003) and Shepherd (2021).

The problem with this physicalist/event-causal picture of agency is that, when causal reality is viewed as nothing but chains of causally related events, everything in the causal world is something that occurs or something that happens. Occurrences and happenings are not things that anyone 'does'. So, when causal reality is viewed as nothing but chains of causally related events, the agent does not seem like an agent anymore, because the agent does not seem to do anything; they seem instead to be merely the setting for events to cause other events. This is the disappearing agent objection, which essentially says that there is something about our concept of agency and something about the idea of the causal world as consisting of nothing but chains of causally related events that don't marry: agency is about agents doing things; a causally related chain of events contains only what occurs or happens. The disappearing agent objection is often dismissed as either begging the question against the physicalist/event-causal account of agency or merely showing that standard physicalist/event-causal accounts need to be modified to include a causal sequence that plausibly plays the functional role of the agent, or only being a problem for libertarian accounts of free will. However, I believe the disappearing agent objection should be taken seriously: there really is a kind of incompatibility between our concept of agency and the idea of the causal world as consisting of nothing but chains of causally related events.

The disappearing agent objection should be taken seriously because the boundary between agential and non-agential does not map onto the divide between event-causal sequences that involve intentional states and those that do not. Sometimes a certain kind of causation by a mental state is what stops an event counting as an instance of agency (deviant causal chain cases); our agency concept extends to cases where agents remain passive and so there is no action to be caused; and our concept of agency extends to cases where there is no mental cause of a bodily movement. What this suggests is that attempting to understand agency in terms of a distinction between event-causal

sequences that involve intentional states and those that do not misconstrues the agency concept.

I concluded that, to properly understand agency, what is needed is a radical departure from the physicalist triad, and in particular the relational approach to causation. Specifically, to understand agency, we need a metaphysical framework that allows us to think of causation as something other than a relation between events. Only then is it possible to see how the causality of action might be something other than a causal relation between mental event and action, and instead something that casts the agent as a causal player, rather than merely the setting for events to cause other events.

In Chapter 6, I outlined a non-relational approach to causation. According to this approach, causation is not always and everywhere a relation, and giving a full account of causation is not merely a matter of explaining what a relation must be like to be a causal relation. Put positively, I maintain that causation can be a process rather than a relation, of which processes like breaking, crushing, bending etc. are more determinate species. My process ontology maintains that processes are universals that substances engage in, and events are instances of processes—they are particular occurrences that come into being when a substance has engaged in a process and completed it.

I argued in Chapter 9 that this non-relational approach to causation, and the process ontology that accompanies it, allows us to put together a more successful understanding of agency. On my view, agents are substances that exercise agential powers, where to exercise a power is for a substance to engage in a process, i.e. for a dynamic state of affairs to obtain. On this view, like other agent-causal accounts of agency, agency is a kind of causation where the agent, who is taken to be a substance, exercises causal power and this exercise of causal power cannot be reduced to causation by an event involving the agent. What makes an action a demonstration of agency is that *the agent* is causing something to happen, where this *causing* of the agent cannot be understood as the causation of one event by another—it is its own special type of causation. However, unlike other agent-causal accounts, I propose that the special type of causation demonstrated in agency is a *process*—not a relation. What it is for a substance to be causing something is for there to be an *activity*—i.e. a way for substances to be effecting change—which the substance is engaging in. Actions are the events that come into existence when agents exercise their agential powers—i.e. engage in processes—and then complete those processes.

I also argued that there are two distinctions crucial to our concept of agential power: the distinction between activity and passivity, and the distinction between one-way and two-way powers. Agency does not reduce to the exercise of active power, because some substances can manifest their agency by remaining passive, and therefore by not engaging in activity. Neither does agency reduce to the exercise of two-way power, because not all substances that cause things to happen do so by exercising two-way powers, but all substances that cause things to happen are agents. My view is that agency is a complex concept

that incorporates both distinctions. Some substances' agential powers are one-way; these substances manifest their agency when they are active but not when they are passive. Other substances' agential powers are two-way; these substances manifest their agency when they are active, but also sometimes when they are passive.

My non-relational approach to causation also opened up new ways of understanding intentional action. Many philosophers have tried to provide an account of intentional action by examining the distinctive sort of explanation with which intentional actions are associated, i.e. rationalising explanations. Davidson (1963) argues that rationalising explanations are causal explanations. They are true if a mental event suitably related to the mental concept cited in the rationalising explanation stands in a causal relation to the action explained. Davidson's argument that rationalising explanations are causal is often taken to justify the claim that mental states or events stand in causal relations to intentional actions. Thus, Davidson's argument is the source of the common view that our conception of ourselves as intentional agents presupposes that mentality is causally relevant in the physical world and that this mental causation should be conceived of in relational terms.

In Chapters 7 and 8, I challenged Davidson's argument that states of desiring and states of believing are causes of the actions they explain. I argued that it is not necessary for an explanation to be causal that its explanandum designate an effect and its explanans designate an item that is the cause of that effect. My non-relational theory of causation implies that facts about causal relations between events are not the only causal facts that causal explanations could answer to. Some causal explanations are made true by the non-relational aspect of causal reality, that is, by facts about substances engaging in processes.

Explanations of intentional action that cite the agent's reasons for acting are the kind of causal explanation that is not made true by causally related events. The most important consideration favouring this view is that it saves two strong intuitions: (a) that reason-giving explanations are causal, and (b) that the mental states cited in reason-giving explanations do not denote items that stand in causal relations to the actions they explain. The second intuition is bolstered by the many arguments offered by non-causalists, which are discussed in Chapter 7, that rationalising explanations need not be considered causal in Davidson's sense to meet Davidson's challenge. The idea that rationalising explanations are causal explanations that answer to the non-relational aspect of causal reality is also supported by the fact that rationalising explanations bear some similarities to both process-citing and disposition-citing explanations.

If these arguments are successful, they show that the fact that we causally explain people's intentional actions by referencing (sometimes directly, sometimes indirectly) their mental states does not justify the contention that, necessarily, whenever there is intentional action there is a causal relation between a mental item and an action or bodily movement. When we say that someone acted intentionally because of what she believed, desired, intended or decided,

these mental concepts need not refer to items that stand in causal relations to physical events. The causal nature of rationalising explanations does not give us any reason to think that there are causal relations between mental items and physical items whenever we act intentionally.

This view, that rationalising explanations are causal explanations that do not designate mental items that stand to the action explained as cause to effect, has consequences for how we ought to think about the nature of intentional action. Most importantly, it casts doubt on the view that intentional actions are distinguished from non-intentional actions by their causes. In Chapter 9, I proposed an alternative view of intentional actions, inspired by Mayr (2011). I proposed that to act intentionally is to engage in a process, and as such is to exercise a power—but a power of a special sort. Intentional actions are manifestations of a special power to organise one's activities into a pattern of determinate form. This power emerges from our possessing two-way powers to act: because we have *many* two-way powers, we also have an extra power to organise our actions into patterns. Rationalising explanations reveal the form of this pattern by attributing mental states with certain contents to the agent. In this way, rationalising explanations are context-placing or structural because they reveal the structure of our activities and make our activities intelligible by helping us see that they are part of a larger pattern of activity. However, rationalising explanations are also disposition-citing because the function of rationalising explanations is to tell us which form the agent was disposed to structure her activities in accordance with.

Is there anything worthy of the name 'mental causation' necessarily on display whenever an agent acts intentionally? I believe we can, and should, answer this question positively. I have mentioned that it is natural to think that some form of mental causation, or 'the reality of causal processes involving cognitive phenomena' as Peter Menzies (2013: 58) puts it, is indispensable to our conception of ourselves as agents who act intentionally and bear moral responsibility. A positive answer to this question is possible once we acknowledge that we need not, and should not, understand 'mental' in 'mental causation' as a 'transferred epithet', as Tim Crane (1995: 219) puts it. Understanding 'mental' in 'mental causation' as qualifying the cause relatum of a causal relation, rather than causation itself, is a prescription of the relational understanding of mental causation.

An alternative conception of the mentality of the causal processes human beings engage in when they act intentionally is that it consists in the fact that these processes are part of a larger pattern of *meaningful*, or *interpretable*, activity.

I have proposed that acting intentionally is to manifest a special power to organise one's activities into a pattern that can be made sense of by appeal to mental concepts. When an agent acts intentionally, the activity the agent is engaging in is part of a larger teleological structure whose form is revealed by attributing knowledge, beliefs, desires or aims to the agent. Furthermore, when you learn that some agent's activity is a manifestation of her desire or an output

of her rational capabilities, you learn that you might be able to alter her activity by altering what she believes about the world, or by changing her desires, usually by reasoning with her, talking to her or persuading her. However, learning this information only makes it the case that you *might* be able to alter the agent's activity. This is because reasoning with an agent in an attempt to prevent them from  $\phi$ ing (or get them to  $\phi$ ) doesn't take away the agent's two-way power to  $\phi$ , so it remains up to her whether she  $\phi$ s or not. Learning about an agent's reasons for acting therefore allows one to manipulate and control the agent's behaviour in a unique way: in a way that leaves the agent's two-way powers intact. I suggest that these are the facts about intentional action that make the causation an agent engages in when they act intentionally count as 'mental'.

Acting intentionally is not mental causation because it consists in actions caused to happen by mental events. Acting intentionally is mental causation in virtue of the fact that the causal activities agents engage in when they act intentionally are part of a larger teleological structure whose form is revealed by attributing mental states to the agent and which we can manipulate in a unique way, i.e. using reasoning and persuasion. For example, the mental causation that is on display when I add salt to the sauce because I think it will make it taste better does not consist in causation of my hand movements by some mental item—e.g. a belief that adding salt will make the sauce taste better. Instead, the causal processes I engage in count as mental causation in virtue of the fact that this particular activity (adding salt to the sauce) is part of a larger pattern of activity whose form and typical trajectory is revealed when it is understood that I want to make the sauce taste better and believe adding salt will achieve that. It is also mental causation in virtue of the fact that persuading me that something else would improve the sauce more effectively is a means by which someone could alter the trajectory of my behaviour while leaving my two-way powers to act intact.

Does my suggestion really capture our intuitive understanding of what mental causation is? Thomas Kroedel (2020) suggests that mental causation can be summarised as the idea that what's going on in your mind makes a difference to what's going on in the world, which is to say that, had our minds been different, our activities would be too. What seems undeniable is that our mental life makes a difference to our bodily life: what we think, what we believe, what we want, what we feel affects what we do with our bodies. It has been a mistake, I think, to interpret this pre-philosophical view as claiming that there is causal interaction between mind and body. The Cartesian notion that our mental life affects what we do with our bodies because we have a mind that causes our body to move is incorrect.

I also think that understanding mental causation as a causal exchange between distinct aspects of ourselves (the mental and the physical) is incorrect. The influence of our mentality on our activities does not reduce to events inside us triggering bodily movements. However, I do not think this is the only way to interpret the naïve idea that what's going on in your mind makes a difference

to what we do with our bodies. Instead, my suggestion is that our minds make a difference to what is going on in the world because we make a difference. When we act intentionally, that is our minds making a difference to the world. Our mental life makes a difference to our bodily life because we have the power to organise our activities into patterns that are made comprehensible by our mental states.

Debates within philosophy of mind tend to centre on which metaphysics of mind best reconciles the claim that mental items stand in causal relations to physical events with plausible principles about what actual causation is like, such as the principle of causal closure. However, if realism about mental causation does not require the relational understanding of mental causation at all, then the problem of mental causation as it is standardly understood may be a pseudo-problem.

Human beings are capable of performing activities that we would naturally describe as ‘mental’, such as imagining and reasoning, and persuading and convincing. Exactly what these activities amount to is a difficult philosophical question. However, it seems to me that these activities are ways to deliberate—individually or in groups—about what beliefs and desires it is best to have, and can be means by which we alter what beliefs or desires an agent has. That we have such capacities is relevant to our bearing moral responsibility.

How it is that we have such capacities is, I think, a very difficult question. How are we able to engage in activities like imagining and reasoning? How does our capacity to imagine, reason, persuade or convince relate to the physical capacities of our bodies? How is it possible that we can change the action plans and projects an agent is disposed to enact by imagining or reasoning or persuading or convincing? I have no idea how to answer these questions. But it is *these* questions—and not questions about how mental items can stand in causal relations to physical events—that constitute the real problem of mental causation. The real mystery is not how mental items can stand in causal relations to physical events but how it is that we can perform mental activities at all.

## References

- Bishop, J 1989 *Natural agency: An essay on the causal theory of action*. Cambridge: Cambridge University Press.
- Brand, M 1984 Intending and acting. *Mind*, 96(381): 121–124.
- Bratman, M 1987 *Intention, plans, and practical reason*. Cambridge, MA: Harvard University Press.
- Crane, T 1995 The mental causation debate. *Aristotelian Society Supplementary Volume*, 69 (1): 211–236. DOI: <https://doi.org/10.1093/aristoteliansupp/69.1.211>
- Dretske, F 1988 *Explaining behavior: Reasons in a world of causes*. Cambridge, MA: MIT Press.

- Enç, B 2003 *How we act: Causes, reasons, and intentions*. New York: Oxford University Press.
- Kroedel, T 2020 *Mental causation: A counterfactual theory*. Cambridge: Cambridge University Press.
- Mayr, E 2011 *Understanding human agency*. New York: Oxford University Press.
- Mele, A 1992 *Springs of action: Understanding intentional behavior*. New York: Oxford University Press.
- Mele, A 2003 *Motivation and agency*. Oxford: Oxford University Press.
- Menzies, P 2013 Mental causation in the physical world. In: Gibb, S C, Lowe, E J and Ingthorsson, R *Mental causation and ontology*. Oxford: Oxford University Press. pp. 58–88.
- Shepherd, J 2021 *The shape of agency: Control, action, skill, knowledge*. Oxford: Oxford University Press.